A 2623-1

American Journal of Public Health

December 1993, Vol. 83, No. 12 Established 1911

Editorials

1662 From Lillian Wald to Hillary Rodham Clinton: What Will Happen to Public Health Nursing? *S. M. Reverby*

1664 Virus Transmission in Health Care Settings: Precautions, Epidemiologic Experience, and Common Sense *J. W. Mosley*

1665 New Lessons from China: Equity and Economics in Rural Health Care *V. W. Sidel*

Annotation

1667 *Salmonella Enteritidis* Infections in France and the United States: Causes vs Causal Models *M. E. Halloran*

Commentary

1670 Hormone Replacement Therapy: The Need for Reconsideration *L. Rosenberg*

Public Health Policy Forum

1674 Editorial: Public Health Nursing—The Opportunity of a Century *M. E. Salmon*

1676 Commentary: Going to the People—Public Health Nursing Today and Tomorrow *J. V. Zerwekh*

Articles

INFECTIOUS DISEASES

1679 An Outbreak of Hepatitis A among Health Care Workers: Risk Factors for Transmission *B. N. Doebbeling, N. Li, and R. P. Wenzel*

1685 Hepatitis B Control in China: Knowledge and Practices among Village Doctors *S. Clayton, H. Yang, J. Guan, Z. Lin, and R. Wang*

1689 Patient-to-Patient Transmission of Hepatitis B in a Dermatology Practice *W. G. Hlady, R. S. Hopkins, T. E. Ogilby, and S. T. Allen*

1694 *Salmonella Enteritidis* Infections in France and the United States: Characterization by a Deterministic Model *L. Watier, S. Richardson, and B. Hubert*

1701 The Health Effects of Swimming at Sydney Beaches *S. J. Corbett, G. L. Rubin, G. K. Curry, D. G. Kleinbaum, and the Sydney Beach Users Study Advisory Group*

1707 Cat Scratch Disease in the United States: An Analysis of Three National Databases *L. A. Jackson, B. A. Perkins, and J. D. Wenger*

1712 Estimates of the US Health Impact of Influenza *K. M. Sullivan, A. S. Monto, and I. M. Longini, Jr*

1717 The Risk of Measles, Mumps, and Varicella among Young Adults: A Serosurvey of US Navy and Marine Corps Recruits *J. P. Struewing, K. C. Hyams, J. E. Tueller, and G. C. Gray*

OTHER TOPICS

1721 Multiethnic Variations in the Pregnancy Outcomes of Military Dependents *G. R. Alexander, G. Baruffi, J. M. Mor, E. C. Kieffer, and T. C. Hulsey*

1726 Custody of Cocaine-Exposed Newborns: Determinants of Discharge Decisions *D. R. Neuspiel, T. M. Zingman, V. H. Templeton, P. DiStabile, and E. Drucker*

1730 Home Health Clients: Characteristics, Outcomes of Care, and Nursing Interventions *K. S. Martin, N. J. Scheet, and M. R. Stegman*

1735 Cause of Death in Patients with End-Stage Renal Disease: Death Certificates vs Registry Reports *T. V. Perneger, M. J. Klag, and P. K. Whelton*

(continued page 1657)

APHA
American Public Health Association

Featuring Infectious Diseases

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

APHA
American Public Health Association

December 1993, Vol. 83, No. 12 ISSN: 0090-0036 Established 1911

Public Health Briefs

INFECTIOUS DISEASES

1739 Infectious Diseases and Mortality among US Nursing Home Residents *C. Beck-Sague, S. Banerjee, and W. R. Jarvis*

1743 Nemesis Revisited: Tuberculosis Infection in a New York City Men's Shelter *E. A. Paul, S. M. Lebowitz, R. E. Moore, C. W. Hoven, B. A. Bennett, and A. Chen*

1746 Antibodies to *Borrelia Burgdorferi* and Tick Salivary Gland Proteins in New Jersey Outdoor Workers *B. S. Schwartz, M. D. Goldstein, and J. E. Childs*

1749 The School-Based Immunization Survey: An Inexpensive Tool for Measuring Vaccine Coverage *L. E. Rodewald, K. J. Roghmann, P. G. Szilagyi, N. L. Winter, J. R. Campbell, and S. G. Humiston*

1751 The Impact of a Public Health Nurse Intervention on Influenza Vaccine Acceptance *M. E. Black, J. Ploeg, S. D. Walter, B. G. Hutchison, E. A. F. Scott, and L. W. Chambers*

1754 Neonatal Tetanus in Peru: Risk Assessment with Modified Enzyme-Linked Immunosorbent Assay and Toxoid Skin Test *L. Vernacchio, G. Madico, M. Verastegui, F. Diaz, T. S. Collins, and R. H. Gilman*

1757 HIV Transmission: Women's Risk from Bisexual Men *R. W. Wood, L. E. Krueger, T. C. Pearlman, and G. Goldbaum*

1759 Acculturation and Gender Differences in Sexual Attitudes and Behaviors: Hispanic vs Non-Hispanic White Unmarried Adults *B. V. Marín, J. M. Tschann, C. A. Gómez, and S. M. Kegeles*

OTHER TOPICS

1762 Inappropriate Halsted Mastectomy and Patient Volume in Italian Hospitals *R. Grilli, F. Mainini, A. Penna, G. Bertolini, N. Scorpiglione, V. Torri, and A. Liberati*

1765 Predictors of Nursing Home Admission in a Biracial Population *M. E. Salive, K. S. Collins, D. J. Foley, and L. K. George*

1768 The Behavioral Risk Factor Surveillance System Questionnaire: Its Reliability in a Statewide Sample *A. D. Stein, R. I. Lederman, and S. Shea*

Health Law and Ethics

1773 Firearms and Health: The Right to Be Armed with Accurate Information about the Second Amendment *J. S. Vernick and S. P. Teret*

Public Health Then and Now

1778 Bringing Care to the People: Lillian Wald's Legacy to Public Health Nursing *K. Buhler-Wilkerson*

Notes from the Field

1787 Burn Prevention through Weatherization Assistance Programs *J. Simon, A. Levensohn, B. B. Metzger, S. Hardman, and S. J. Klein*

1788 The Northwest Center for Public Health Practice *G. S. Omenn, M. Oberle, J. Gale, J. J. Hoover, D. Sandlin, and J. Tapp*

Other Journal Departments

1658 APHA Masthead

1660 What *AJPH* Authors Should Know

1661 Advertisers' Index

1662 AJPH Masthead

1790 Letters to the Editor

1795 Book Corner

1799 Annual Meeting Audiocassette Order Form

1805 Job Opportunities

1815 Volume Index

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 Fifteenth Street, NW, Washington, DC 20005; (202) 789-5600. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS
Helen Rodriguez-Trias, MD, *President*
Joyce C. Lashof, MD, *Immediate Past President*
Eugene Feingold, PhD, JD, *President-Elect*
G. Ted Fairchild, RD, MPH, *Vice-President (USA)*
Robert Evans, PhD, *Vice-President (Canada)*
Julio Teja Perez, MD, *Vice-President (Latin America)*
Jay H. Glasser, PhD, *Treasurer*
Barry S. Levy, MD, MPH, *Chair, Executive Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*
Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD
Barry S. Levy, MD, MPH (1993), *Chair*
Marilyn Aguirre-Molina, EdD, MS (1993), *Vice-Chair*
Eugene Feingold, PhD, JD, *President-Elect*
Jay H. Glasser, PhD, *Treasurer*
Janet Heinrich, DrPH, MPH (1993)
Alan R. Hinman, MD, MPH (1995)
Margo D. Kerrigan, MPH (1995)
Joyce C. Lashof, MD, *Immediate Past President*
Joanne Lukomnik, MD (1996)
Linda Ray Murray, MD, MPH (1996)
Steven Uranga McKane, DMD, MPH (1995)
Hila Richardson, DrPH, MPH (1996)
Helen Rodriguez-Trias, MD, *President*
Allan G. Rosenfield, MD (1994)
Reed V. Tuckson, MD (1994)
Vacancy (1994)

EX OFFICIO
Fernando M. Treviño, PhD, MPH, *Executive Director*
Frances Kissling, *Chair, Action Board*
George E. Pickett, MD, MPH, *Speaker, Program Development Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*

American Journal of Public Health

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, Inc, 1015 Fifteenth St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 Fifteenth St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 Fifteenth St, NW, Washington, DC 20005. AJPH allotted $40; TNH allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 Fifteenth St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, AJPH Reprints, APHA, 1015 Fifteenth St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1993 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The American Journal of Public Health is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

An Outbreak of Hepatitis A among Health Care Workers: Risk Factors for Transmission

ABSTRACT

Objectives. The purpose of this study was to investigate a nosocomial outbreak of hepatitis A that occurred in the burn treatment center of a referral hospital.

Methods. Retrospective cohort and case-control studies were performed to determine acquisition rates and risk factors for transmission. Adjusted infection rates were calculated by week of exposure. A case-control study was conducted to determine potential mechanisms for nosocomial acquisition. Recently infected health care workers were defined as case patients; exposed, serosusceptible health care workers without infection served as controls.

Results. The outbreak of hepatitis A affected 11 health care workers and 1 other burn patient (1 secondary patient case). All 11 health care workers became ill after the admission of a man and his 8-month-old son who developed hepatitis A while in the hospital. The cumulative incidence risk ratio was elevated for health care workers caring for either the infant or the father during the same week of exposure. The case-control study implicated the behavior of eating on the hospital ward as the single most important risk factor for infection.

Conclusion. Inadequate handwashing and subsequent oral contamination appear responsible for the outbreak. Hospitals may witness other institutional outbreaks if health care workers regularly eat on the wards. (*Am J Public Health.* 1993;83: 1679–1684)

Bradley N. Doebbeling, MD, MS, Ning Li, MB, MS, and Richard P. Wenzel, MD, MSc

Introduction

Nosocomial outbreaks of hepatitis A are relatively uncommon. Several food-related epidemics have been reported.[1,2] Most nosocomial outbreaks have involved exposure either to an asymptomatic infant or to a young child, typically after the transfusion of blood from an asymptomatic donor.[3–8] Occasionally, outbreak clusters have been reported after contact with an older child or adult with vomiting, diarrhea, or fecal incontinence.[9–14] Several large outbreaks in neonatal intensive-care units have been reported recently.[4,5,7,8] Despite extensive investigations, most previous reports failed to identify significant risk factors for nosocomial transmission, other than associated care of the infected individual. Rosenblum and colleagues[8] recently reported an outbreak of hepatitis A in a neonatal intensive-care unit, implicating the behaviors of drinking beverages in the unit and failure to wear gloves routinely as risk factors for transmission to health care workers.

We report here the first nosocomial outbreak of hepatitis A in a burn treatment center. The outbreak occurred between January and March 1990 in a tertiary-care hospital. A 32-year-old man and his 8-month-old son were admitted to the burn treatment center on December 9, 1989, after each had sustained large body surface area burns in a house fire, 70% and 40%, respectively. On January 8, 1990, the father became jaundiced, and both were found to have acute hepatitis A infections. Despite institution of appropriate isolation precautions,[15] 11 health care workers and 1 other burn patient became clinically ill with hepatitis A. We undertook an epidemiologic investigation to identify important risk factors for transmission in the hospital. Our investigation is the subject of this report.

Background

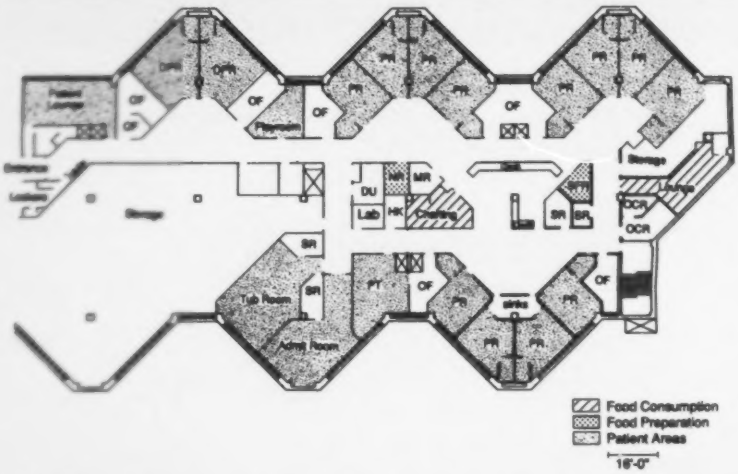
The hospital is a 900-bed referral center with a 16-bed burn treatment center. The typical nursing staff to patient ratio on the burn treatment center is approximately 1:3, although it varies depending on the acuity of the patient and the time of day. Health care workers regularly assigned to the burn treatment center often eat in the staff lounge or in a charting room, located approximately 15–50 feet from most of the patient rooms (Figure 1). Staff food is stored in a refrigerator in a designated staff food room, the staff lounge, or occasionally in the patient nourishment room. Infants and young children with burns are allowed to play in a playpen and eat beside the desks in the center of the burn treatment area. Additionally, outpatients with burns are routinely seen in follow-up on the ward and often bring in snack foods to share with the burn treatment personnel. Two sinks are located at the entrance to the center, and hand-washing is strongly encouraged

Bradley N. Doebbeling and Richard P. Wenzel are with the Department of Internal Medicine, University of Iowa College of Medicine, Iowa City, Iowa. Richard P. Wenzel is also with the Program of Epidemiology, The University of Iowa Hospitals and Clinics. Ning Li is with the Department of Preventive Medicine and Environmental Health, Division of Biostatistics, The University of Iowa College of Medicine.

Requests for reprints should be sent to Bradley N. Doebbeling, MD, MS, Department of Internal Medicine, C41L GH, The University of Iowa College of Medicine, 200 Hawkins Dr, Iowa City, IA 52242-1081.

This paper was accepted April 26, 1993.

Editor's Note. See related editorial by Mosley (p 1664) in this issue.



Note. Staff members consumed food (hatched pattern) in the staff lounge, charting room, and occasionally in individual offices (OF). Food storage and preparation areas (dotted pattern) for the staff members included the staff food room (SFR), patient nourishment room (PNR), and staff lounge. Patient care areas (shaded) include sinks in each single and double patient room (PR and DPR, respectively). Other abbreviations include dirty utility (DU) room, medication room (MR), housekeeping (HK), physical therapy (PT), supply rooms (SR), and on-call rooms (OCR).

FIGURE 1—Floor plan of the burn treatment center.

on entering and leaving the ward with a display of prominent signs.

Methods

Epidemic Investigation

Selected burn treatment personnel and the burn patient with the secondary case of hepatitis were interviewed to determine potential common exposures and risk factors. The burn treatment center's charting room, staff lounge, refrigerators, sinks, ice machine, nourishment room, hydrotherapy room, rest rooms, water supply, and all patient care areas were inspected by members of the hospital epidemiology staff.

Screening. All health care workers assigned to the center in the prior 6 weeks (n = 154), all inpatients at the time of identification of the outbreak on February 9, 1990 (n = 12), and all patients discharged in the previous month (n = 45) were contacted and offered the option of screening with liver function tests and a serologic test for hepatitis A virus immunoglobulin M antibody (IgM anti-HAV test) (enzyme immunoassay, Abbott Laboratories, Chicago, Ill). Additionally, all food preparation workers in the hospital and the dietary services staff working on the burn center (n = 119) were contacted and similarly screened. All exposed employees of the burn treatment center who were IgM anti-HAV negative were tested for total anti-HAV antibody (anti-HAV, Abbott). A repeat IgM anti-HAV test was obtained for all consenting exposed health care workers regularly assigned to the burn treatment center 6 weeks later (n = 37) to determine whether seroconversion had occurred despite prophylaxis.

Case definition. A health care worker was defined as exposed if he or she had direct patient contact on the burn center in the 6 weeks (December 28, 1989, to February 7, 1990) before the outbreak. A preliminary case definition required elevation of serum levels of one of the liver transaminases to twice the upper limit of normal or symptoms consistent with acute HAV infection. Final inclusion as a case subject required a positive IgM anti-HAV titer, with or without clinical symptoms. A susceptible control subject was an exposed health care worker who tested negative for anti-HAV and who was in one of the same occupational categories as the health-care-worker case subjects (n = 60). Individuals with incomplete test results or prior immunity were excluded from further analysis.

Immunoprophylaxis. All asymptomatic health care workers in the burn treatment center who had contact with the index cases, all current patients, and all patients discharged in the previous month were contacted and offered immunoglobulin immunoprophylaxis. Health care workers were informed that the efficacy of immunoglobulin immunoprophylaxis has been demonstrated only when it is given within 2 weeks of exposure. Additionally, regional exchange nurses temporarily assigned to the burn treatment center to cover staffing shortages and all new patients admitted for the subsequent 6 weeks were offered immunoglobulin. All close contacts of those persons who were clinically ill or who had elevated liver transaminases were contacted to be screened and offered immunoglobulin. Consenting exposed individuals were given a single dose of immunoglobulin 0.02 ml/kg (Gamastan, Cutter Biologicals, West Haven, Conn) intramuscularly in the deltoid muscle. All case subjects were contacted by telephone 4 months later to determine whether any symptomatic secondary cases had occurred among their family members or close contacts.

Four burn treatment center nurses had been "pulled" to staff four other intensive-care units in the 48 hours before they became clinically ill. The patients for whom they cared and health care workers in close contact with the infected nurses were contacted and offered immunoglobulin prophylaxis.

Index cases investigation. The source of the initial index cases was investigated by interviewing the adult index case patient and his family members regarding potential risk factors and any possible common sources of exposure to hepatitis A before admission to the hospital. The Iowa State Health Department inspected a single family day-care home that cared for the index infant and six other children before the outbreak. Serum was obtained for an anti-HAV test from five of the six children, the babysitter, and family members of the babysitter. However, a sample could not be obtained from one of the children. Additionally, blood bank records were reviewed to determine the individuals donating blood to the two index case patients. None of the donors had given blood to both patients; all were contacted and asked to report any clinical illness in the 2 months before their donations.

Patient care study. A retrospective cohort study of health care workers was performed to evaluate the risk of infection associated with caring for the patients with the index cases. Nursing care assignments for the 6 weeks before the period of the peak incidence were examined to determine which nursing personnel had been

exposed to the index case patients. The study population included nurses and nursing students working in the burn treatment center during the 6-week period. Serosusceptible nurses caring for either of the index case patients were classified as exposed if assignment to one of the index patients occurred at least 2 weeks before the development of illness by the health care worker. Adjusted infection rates by week of exposure were calculated and were stratified by the number of shifts worked with either or both of the patients. Cumulative incidence ratios (relative risk ratios) were also calculated by week of exposure to either index patient.

Risk factor study. A case–control study was performed to evaluate behavioral risk factors and common exposures among health care workers for development of hepatitis A infection. A 50-item self-administered questionnaire was developed to assess potential risk factors for hepatitis A infection such as patient care practices (e.g., hand-washing, glove use, providing hydrotherapy, specific types of care of infants and adults), personal habits while working (e.g., eating, drinking, smoking, biting fingernails), any group exposures (e.g., potlucks, shared meals, dining out), environmental exposures, and specific food items eaten while at work. The questionnaire was pretested, revised, and then mailed to all 154 health care workers assigned to work in the burn treatment center in the 6 weeks before the outbreak's peak. All questionnaires were identified only by a confidential study number known to the investigators. Three sequential mailings of the questionnaire, which were followed by telephone contacts, were used to improve the response rate.

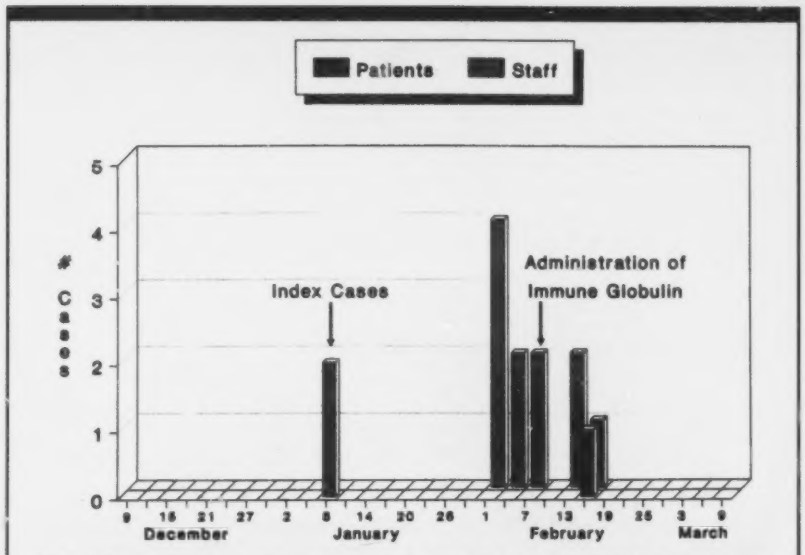
Statistical Methods

The statistical analyses included the use of the Mantel-Haenszel chi-square statistic, Fisher's two-tailed exact test, and odds ratios with 95% confidence intervals for all significance testing.[16] For the tables with zero entries, 0.5 was added to each cell; the Mantel-Haenszel estimator of the odds ratio and the test-based confidence interval were determined. Stepwise logistic regression was conducted by using SAS[17] to evaluate all variables that were identified by the univariate approach ($P < .10$). The significance level was set at .05. All P values reported are two-tailed.

Results

The IgM anti-HAV test was positive in the adult index case patient 1 month after admission. His asymptomatic 8-month-old son was also IgM anti-HAV positive at the same time. Serologic tests for hepatitis A (IgM-specific and total anti-HAV antibody) and liver function tests for both patients had been negative on admission. At the time of the investigation, no additional blood samples were available for antibody testing from the first month of the patients' hospitalization. Neither patient had vomiting or diarrhea. Inspection of the burn treatment center revealed sinks and dispensers for hand-cleansing agents located in each room, and all were found to be in good repair. Boxes of gloves and gowns and masks were readily available in each patient's room. All burn treatment personnel were asked to refrain from eating or drinking on the ward after identification of the outbreak.

A total of 154 health care workers were identified as having been exposed to patients on the burn treatment center in the 6 weeks before the health care workers became ill. Serologic tests for anti-HAV antibody in all consenting exposed workers demonstrated that only 8.7% of the 126 workers screened had detectable immunoglobulin G antibody to the hepatitis A virus. Three health care workers with clinical hepatitis A infection and the secondary case patient were initially seronegative for IgM anti-HAV antibody despite symptoms and elevated hepatic transaminases, but subsequently became seropositive in 5 to 7 days.

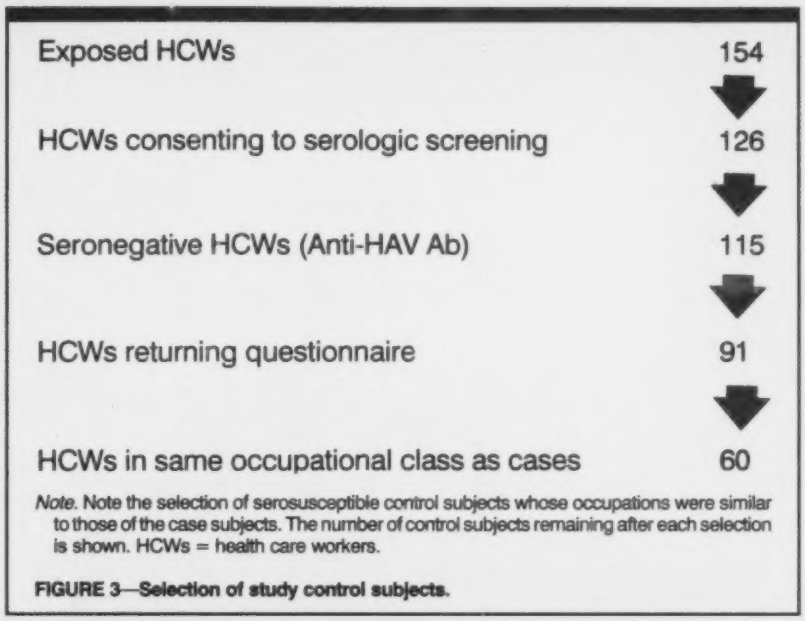


Note. Although the exact date of onset in the infant is uncertain, both index cases are represented at the date of onset of clinical illness in the adult. The single patient represented by the dark bar in February signifies the only secondary patient case. Not shown are the asymptomatic mother of the infant and an asymptomatic household contact of one of the nurse cases whose date of seroconversion is unclear. Note the incubation period of approximately 1 month among the health care workers.

FIGURE 2—Epidemic curve of an outbreak of hepatitis A on a burn treatment center.

Eventually a total of 11 health care workers developed IgM anti-HAV antibody, and all became clinically ill (Figure 2). Two required hospitalization for dehydration. The infection rates among serologically defined susceptible individuals ranged from 14% (7 of 50) among staff nurses and nursing assistants to 33% (2 of 6) among ward clerks and 67% (2 of 3) among nursing students. All case subjects had contact with the infant, whereas only 8 of 11 case subjects (73%) also had contact with the father. The 32-year-old mother of the index infant was also seropositive for IgM anti-HAV antibody, but remained asymptomatic. No symptomatic secondary cases occurred in close contacts or family members of the infected case subjects. The mother of the child was seropositive when checked in early February, although she was asymptomatic. One asymptomatic seroconversion occurred after prophylaxis with immunoglobulin in the fiancé of one of the nurse case subjects. None of 37 seronegative health care workers whose serology was rechecked at 4 to 6 weeks developed anti-HAV antibody.

Importantly, four cases of hepatitis A occurred in the county of residence of the infant and his father during November 1989, followed by 8 and 17 cases, respec-



Note. Note the selection of serosusceptible control subjects whose occupations were similar to those of the case subjects. The number of control subjects remaining after each selection is shown. HCWs = health care workers.

FIGURE 3—Selection of study control subjects.

TABLE 1—Hepatitis A among Health Care Workers in a Burn Treatment Center, January–February 1990: A Case–Control Questionnaire Study on Risk Factors

Exposure Variable[a]	Case Subjects (n = 11) Yes	Case Subjects (n = 11) No	Control Subjects (n = 60) Yes	Control Subjects (n = 60) No	Odds Ratio	95% CI[b]
Eating food on the burn center (BC) (frequently vs occasionally or never)	10	1	24	34	14.2	2.49, 80.50
Bringing in snacks to share in prior 6 weeks	7	4	10	50	8.8	2.44, 31.34
Eating snacks brought in by coworkers	10	1	36	24	6.7	0.99, 44.48
Close contact with mother of infant	2	9	2	58	6.4	0.99, 41.95
Providing hydrotherapy	9	2	26	34	5.9	1.33, 26.13
Spending ≥3 hours daily with BC patients	9	2	27	33	5.5	1.23, 24.62
Working in the tub room	9	2	29	31	4.8	1.06, 21.88
Care of infants: playing[c]	11	0	52	8	3.7	0.20, 69.23
Eating food from the common refrigerator	9	2	32	28	3.9	0.85, 18.34
Drinking coffee prepared on the BC	8	3	26	34	3.5	0.88, 13.78
Care of infants: holding/carrying[c]	11	0	53	7	3.2	0.17, 60.55
Eating at a BC potluck	3	8	9	51	2.1	0.48, 9.43
Eating out with coworkers in prior 6 weeks	4	7	14	44	1.8	0.37, 8.39
Chewing on pens or pencils	4	7	15	45	1.7	0.44, 6.68
Eating cookies made by index patients' family	3	8	13	47	1.4	0.31, 5.89
Care of infants: physical therapy	8	3	41	19	1.2	0.29, 5.23
Smoking	3	8	13	47	1.4	0.31, 5.89
Care of infants: dressing changes	9	2	48	12	1.3	0.21, 5.97
Eating cafeteria food on the BC	10	1	56	4	0.7	0.07, 7.12
Biting fingernails	1	10	16	44	0.3	0.04, 2.42

Note. All of the control subjects did not respond to each questionnaire item, so the total number of health care workers who did not become ill does not equal 60 for all items.
[a]All variables have a *P* value of <.10 or were suggested by previous studies as potentially important.
[b]95% confidence interval about the odds ratio was calculated with test-based Mantel-Haenszel estimates about the odds ratio.
[c]Logit estimators of the 95% confidence interval use a correction of 0.5 to every cell of those tables that contain a zero.

tively, during the next 2 months. No other clustering of hepatitis A infections was noted at restaurants or other facilities in the county of origin during the period of the outbreak. Anti-HAV serology was obtained from the two donors reporting symptomatic illness coincident with the blood donations.

Characteristics of the Study Samples

Ninety-two percent (n = 132) of the questionnaires from potentially eligible controls (n = 154) were returned; all case subjects returned the questionnaire. Controls were required to be both serosusceptible (negative for total anti-HAV antibody) and to have returned the questionnaire (n = 91, Figure 3). The control group was further refined by limiting the group to the same job classifications (i.e., those with maximal patient exposure) as those of the health-care-worker case subjects (e.g., nurses, nursing assistants, nursing students, and ward clerks; n = 60). The reported rates of hand-washing after patient care, hand-washing after diaper changing, and wearing gloves when contact was anticipated with blood and body fluids or secretions were similar for both groups. After patient care, 82% of the case patients and 75% of the controls reported that they washed their hands routinely (relative risk [RR] = 1.09, 95% confidence interval [95% CI] = 0.77, 1.55). After diaper changing, 91% of the case patients and 100% of the controls reported always washing their hands (RR = 0.95, 95% CI = 0.85, 1.06). Similarly, 82% of the case and 63% of the control subjects reported consistent use of gloves when contact with blood or body fluids was anticipated (RR = 1.29, 95% CI = 0.85, 1.98).

Nursing assignments. The infection rates for nurses assigned to either patient were the highest during the exposure week of January 2 through January 8, 1990, just before the onset of symptoms in the father. During that week, the infection rates among nurses assigned to two or more shifts with the infant were higher than those for nurses assigned to a single shift (4 of 5 [80%] vs 1 of 6 [17%], respectively). Similarly, infection rates were higher for nurses assigned to two or more shifts with the infant's father than for nurses assigned to a single shift (3 of 3 [100%] vs 1 of 13 [8%], respectively). The cumulative incidence risk ratio was elevated for nurses caring for either th infant (RR = 14.6, 95% CI = 3.4, 63.2) or his father (RR = 10.6, 95% CI = 2.2, 50.2) during that same week. The only two nurses assigned to care for both patients during the same week also became ill.

Risk factor assessment. The univariate analysis (Table 1) identified the following significant risk factors for hepatitis A infection: eating on the ward (odds ratio [OR] = 14.2, 95% CI = 2.5, 80.5), bringing snack foods to share with others (OR = 8.8, 95%

CI = 2.4, 31.3), and providing hydrotherapy treatments (OR = 5.9, 95% CI = 1.3, 26.1). Of 12 different types of patient care given to adults or infants, none appeared to predispose workers to hepatitis A infection. No other risk factors (e.g., smoking, chewing on pens or pencils, eating specific snack foods at work, eating out with coworkers) were statistically significant risk factors. After correction for confounding variables with a stepwise logistic regression analysis, the single important risk factor for the development of infection was frequent eating on the ward (OR = 14.2, $P = .0029$). No additional variables met the criteria for retention in the model ($P < .05$) or improved the single variable model.

Discussion

Health care workers caring for thermally injured patients are potentially exposed to a variety of blood-borne pathogens daily through the open burn wounds that must be debrided with hydrotherapy and frequent dressing changes. In our burn treatment center, both case and control subjects reported a high level of prior compliance with routine handwashing, glove use, and other barrier forms of universal precautions. Nevertheless, an important outbreak of hepatitis A occurred in a burn treatment center involving 11 health care workers and 1 patient with a secondary case. Once the epidemic was recognized, reemphasis of good hygienic and isolation practices and the administration of immunoglobulin apparently prevented other secondary cases. Although the relative contributions of each of these measures in terminating the outbreak cannot be assessed in the current study, we would recommend the use of a similar approach for control of a nosocomial outbreak.

We were unable to determine the source of hepatitis A infection in the two index patients. The usual incubation period of 10–50 days (mean = 1 month) is consistent with either nosocomial or community-acquired disease.[18] The elevated cumulative incidence ratios among nurses exposed to either case in the week before onset of symptoms in the adult index patient suggest that both index patients were most infectious during that period, which corresponds to the time of greatest fecal shedding of hepatitis A virus. We also have epidemiologic evidence of an exposure dose–response relationship: the infection rates were higher for nurses assigned to more than one shift than for nurses assigned to only one shift with either patient and highest among nurses assigned to care for both patients. However, the infant son may have been infected before his arrival at the hospital, explaining the exposure to and timing of the disease in his father. Infants usually remain asymptomatic, liver enzymes may not be elevated, and secretion of virus in the stool may continue for weeks.[8]

Similar exposure dose–response relationships have been reported in the literature. In one multistate outbreak of 55 cases, the 16% infection rate among susceptible nurses was four times higher than that among physicians with considerably less exposure to case patients.[4] Moreover, nurses working 12-hour shifts were significantly more likely to become infected than those working 8-hour shifts. In another epidemic, nurses assigned to two or more shifts with an infected infant with an ileostomy had an infection rate 4.6 times higher than those assigned to a single shift.[5]

Because the index patients did not have direct contact with each other until after the onset of illness in the father, a nosocomial source for their infections appears unlikely. Nevertheless, a possible nosocomial source would be a contaminated blood product given to both patients. Transfusion-associated hepatitis in neonates has been recognized as the source of several nosocomial outbreaks involving health care workers.[3-6] However, in our center, none of the donors gave blood to both patients. Moreover, of the 33 donors contributing blood to either patient, only 2 reported feeling ill near the time of their donation. Hepatitis A serologic tests from both of the latter donors were negative.

The outbreak occurred in the setting of a national increase in the incidence of hepatitis A in the United States.[19] At the time of the outbreak, the incidence of community-acquired hepatitis A in the state of Iowa was fivefold higher than that in the previous year, with one fourth of the cases occurring in the same county of origin as the index cases.

Up to one third of community-acquired cases of hepatitis A have been linked to day-care centers.[20-22] However, in our study, none of the children tested or family members of the babysitter from the small day-care home that had previously cared for the index infant were seropositive. Nevertheless, one of the six children cared for in the center could not be contacted for testing. Similarly, no apparent common source of exposure could be identified for the index family in the 2 months before admission.

A case–control study using seronegative, susceptible health care workers as controls convincingly demonstrated that frequent eating on the ward was the single most important risk factor for development of hepatitis A infection. Importantly, previous investigations of nosocomial outbreaks of hepatitis A have typically failed to implicate a particular mode of transmission. The finding that hepatitis A infection is clearly associated with eating on the hospital ward is not only biologically plausible, but also is entirely consistent with the acknowledged fecal-oral route of transmission of hepatitis A infections.

Our data indicating nonroutine handwashing by approximately 25% of health care workers and nonroutine use of gloves by 37% of health care workers are self-reported, retrospective, and conservative estimates of compliance with standard hygienic procedures. These results support the previous findings of Rosenblum et al.,[8] who found a significantly increased risk of infection during an outbreak in a neonatal intensive-care unit among health care workers who did not wear gloves when taping an intravenous line.

We speculate that hygiene was suboptimal in the burn center before the outbreak and that hand contamination probably occurred. Hand contamination may have occurred after providing hydrotherapy (9 of 11 cases) or after other patient care activities. Subsequently, nosocomial infection likely occurred either through eating snacks or sharing food on the ward. The strategic placement of sinks at the entrance to the center may have resulted in improved hand-washing on leaving the burn treatment center and explain the lack of association of illness with eating in the hospital cafeteria. A food-related outbreak was also unlikely because there was geographic clustering of the cases on a single patient unit, serologic tests for hepatitis A among dietary workers were negative, and there was no supportive evidence in the case–control study. Importantly, none of the snack foods available on the center or the specific food items at the potluck dinner were associated with illness. However, we are unable to exclude absolutely the possibility that food contamination rather than, or in addition to, contamination of health care workers' hands was the apparent mode of transmission.

Previously reported outbreaks have been linked to patients with vomiting, diarrhea, or fecal incontinence.[9-14] An out-

break of hepatitis A in a pediatric intensive-care unit in 1981 involving 10 secondary cases suggested that food was a potentially important risk factor.[9] The apparent index patient in that outbreak had been incontinent of feces for several days before his death. His family had frequently brought in food that they shared with the boy and members of the staff. Sharing food with patients or their families, drinking coffee at work, sharing cigarettes, and eating in the nurses' office increased the likelihood of infection, although none of the variables reached statistical significance.[9] Moreover, in another study, case patients were almost three times more likely to smoke than were controls (36% vs 13%), suggesting the importance of the fecal-oral route of transmission.[5] Similarly, Rosenblum and colleagues[8] reported that drinking beverages in the intensive-care unit was a significant risk factor for infection.

Recent data demonstrated that hepatitis A virus survives well on the fingerpads, despite drying, for 4 hours.[23] Additionally, the virus was transferred effectively between fingerpads and from fingerpads to clean surfaces despite prolonged drying. After suspensions of the hepatitis A virus were dried for 20 minutes, an inoculum of the virus was transferred from an inanimate surface to the fingerpads. Both pressure and friction significantly increased the transfer of hepatitis A virus. The results of this experimental study support the importance of human hands in the direct spread of hepatitis A observed in the current outbreak.

The diagnosis of acute hepatitis A is usually confirmed easily by the presence of a positive IgM anti-HAV serologic test, which is typically positive very early in cases of acute infection.[18,24] Four of our patients developed nausea, vomiting, and elevated alanine aminotransferase levels but were initially seronegative. Each of these case patients, however, developed IgM anti-HAV antibody over the next 5 to 7 days. Our findings confirm those of Zachoval and colleagues,[25] who recently reported four patients with early acute hepatitis A infection who were initially seronegative for IgM anti-HAV antibody. Therefore, patients suspected of having early hepatitis A infection but who have a negative IgM anti-HAV test should have the test repeated several days later.[25]

Our results confirm the empiric observations of Drusin and colleagues[9] and Reed et al.[10] and are consistent with previous guidelines[26] suggesting that eating on the hospital ward when hygiene is suboptimal predisposes health care workers and patients to hepatitis A. Moreover, the combination of a low level of population immunity and an increase in the incidence of hepatitis A infection in the United States suggests that other hospitals may soon experience similar clusters of infection if health care workers continue to eat on the hospital ward. □

Acknowledgments

This paper was presented in part at the 30th Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC), Atlanta, Ga, in October 1991 (abstract 1022).

We thank Jeanne Lyons and Jackie Omaher for their assistance and Drs Jack Stapleton, Douglas LaBrecque, and Loreen Herwaldt for their advice during the investigation of the outbreak.

References

1. Eisenstein AB, Aach RD, Jacobsohn W, Goldman A. An epidemic of infectious hepatitis in a general hospital: probable transmission by contaminated orange juice. *JAMA.* 1963;185:171–174.
2. Meyers JD, Romm FJ, Tihen WS, Bryan JA. Food-borne hepatitis A in a general hospital: epidemiologic study of an outbreak attributed to sandwiches. *JAMA.* 1975;231: 1049–1053.
3. Seeberg S, Brandberg A, Hermodsson S, Larsson P, Lundgren S. Hospital outbreak of hepatitis A secondary to blood exchange in a baby. *Lancet.* 1981;1:1155–1156. Letter.
4. Noble RC, Kane MA, Reeves SA, Roeckel I. Posttransfusion hepatitis A in a neonatal intensive care unit. *JAMA.* 1984;252:2711–2715.
5. Azimi PH, Roberto RR, Guralnik J, et al. Transfusion-acquired hepatitis A in a premature infant with secondary nosocomial spread in an intensive care nursery. *Am J Dis Child.* 1986;140:23–27.
6. Giacoia GP, Kasprisin DO. Transfusion-acquired hepatitis A. *South Med J.* 1989;82: 1357–1360.
7. Klein BS, Michaels JA, Rytel MW, Berg KG, Davis JP. Nosocomial hepatitis A: a multinursery outbreak in Wisconsin. *JAMA.* 1984;252:2716–2721.
8. Rosenblum LS, Villarino ME, Nainan OV, et al. Hepatitis A outbreak in a neonatal intensive care unit: risk factors for transmission and evidence of prolonged viral excretion among preterm infants. *J Infect Dis.* 1991;164:476–482.
9. Drusin LM, Sohmer M, Groshen SL, Spiritos MD, Senterfit LB, Christenson WN. Nosocomial hepatitis A infection in a paediatric intensive care unit. *Arch Dis Child.* 1987;62:690–695.
10. Reed CM, Gustafson TL, Siegel J, Duer P. Nosocomial transmission of hepatitis A from a hospital-acquired case. *Pediatr Infect Dis.* 1984;3:300–303.
11. Krober MS, Bass JW, Brown JD, Lemon SM, Rupert KJ. Hospital outbreak of hepatitis A: risk factors for spread. *Pediatr Infect Dis.* 1984;3:296–299.
12. Orenstein WA, Wu E, Wilkins J, et al. Hospital-acquired hepatitis A: report of an outbreak. *Pediatrics.* 1981;67:494–497.
13. Edgar WM, Campbell AD. Nosocomial infection with hepatitis A. *J Infect.* 1985;10: 43–47.
14. Goodman RA, Carder CC, Allen JR, Orenstein WA, Finton RJ. Nosocomial hepatitis A transmission by an adult patient with diarrhea. *Am J Med.* 1982;73:220–226.
15. American Public Health Association. Viral hepatitis A. In: Benenson AS, ed. *Control of Communicable Diseases in Man.* 15th ed. Washington, DC: American Public Health Association; 1990:197–200.
16. Schlesselman JJ. *Case-Control Studies.* New York: Oxford University Press; 1982: 177–211.
17. SAS Institute Inc. *SUGI Supplemental Library User's Guide, Version 5.* Cary, NC: SAS Institute; 1986.
18. Hollinger FB, Glombicki AP. Hepatitis A virus. In: Mandell GL, Douglas RG Jr, Bennett JE, eds. *Principles and Practice of Infectious Diseases.* New York: Churchill Livingstone; 1990:1383–1399.
19. Centers for Disease Control. Foodborne Hepatitis A—Alaska, Florida, North Carolina, Washington. *MMWR.* 1990;39:228–232.
20. Benenson MW, Takafuji ET, Bancroft WH, Lemon SM, Callahan MC, Leach DA. A military community outbreak of hepatitis type A related to transmission in a child care facility. *Am J Epidemiol.* 1980; 112:471–481.
21. Hadler SC, Webster HM, Erben JJ, Swanson JE, Maynard JE. Hepatitis A in day-care centers. A community-wide assessment. *N Engl J Med.* 1980;302:1222–1227.
22. Vernon AA, Schable C, Francis D. A large outbreak of hepatitis A in a day-care center: association with non-toilet-trained children and persistence of IgM antibody to hepatitis A virus. *Am J Epidemiol.* 1982; 115:325–331.
23. Mbithi JN, Springthorpe S, Boulet JR, Sattar SA. Survival of hepatitis A virus on human hands and its transfer on contact with animate and inanimate surfaces. *J Clin Microbiol.* 1992;30:757–763.
24. Froesner GG, Deinhardt F, Roggendorf M, Scheid R, Bayerl B, Zachoval R. Laboratory diagnosis of hepatitis A infection. In: Bianchi L, Gerok W, Sickinger K, Stalder GA, eds. *Virus and the Liver.* Lancaster, England: MTP Press; 1980:17–26.
25. Zachoval R, Kroener M, Brommer M, Deinhardt F. Serology and interferon production during the early phase of acute hepatitis A. *J Infect Dis.* 1990;161:353–354. Letter.
26. Doebbeling BN, Wenzel RP. Nosocomial Viral Hepatitis. In: Mandell GL, Douglas RG Jr, Bennett JE, eds. *Principles and Practice of Infectious Diseases.* 3rd ed. New York: Churchill Livingstone; 1990: 2215–2221.

Hepatitis B Control in China: Knowledge and Practices among Village Doctors

ABSTRACT

Objectives. To guide development of a hepatitis B control program in China, we investigated village doctors' sterilization practices, injection use, vaccination levels, and knowledge and counseling skills.

Methods. The head doctor from each village health station in a rural county (n = 260) completed a self-administered questionnaire.

Results. Of the respondents, 94.2% reported adequate sterilization of needles; however, inadequate sterilization of syringes and acupuncture needles was common. Injections were found to be frequent and profitable. Of head doctors, 16.2% reported that most infants in their villages are vaccinated against hepatitis B. There was very high knowledge of modes of hepatitis B transmission, but it was not used to formulate advice for a hypothetical carrier. Knowledge of the consequences, prevalence, and use of condoms to prevent sexual transmission was lower.

Conclusions. The sterilization of acupuncture needles, reuse of syringes, and frequent injections are practices that may transmit the hepatitis B virus. Research into issues of vaccine acceptance is needed to expand immunization. A review of treatment practices and better training in counseling for village doctors is suggested. (*Am J Public Health.* 1993;83:1685–1688)

Serena Clayton, MA, Honglian Yang, Jihui Guan, MB, Zhongsu Lin, and Rui Wang

Introduction

In the People's Republic of China, approximately 10% of the population is chronically infected with the hepatitis B virus (HBV),[1] with rates as high as 15% in the southern coastal provinces such as Fujian.[2] Approximately 25% of individuals who are chronic carriers of HBV will die of cirrhosis or primary hepatocellular carcinoma.[3] In Fujian, liver cancer accounts for 22.36% of cancer mortality in males and 13.75% in females.[4] HBV is transmitted primarily percutaneously through blood, blood products, and semen. Perinatal transmission has been estimated to account for 30% to 50% of HBV carriage in China.[5] The spread of HBV through medical practices such as injections[6] and acupuncture remains an important concern, as does intrafamilial transmission. Because the likelihood of becoming a chronic carrier decreases as age at infection increases,[7] prevention of infection in infancy is the top priority. The hepatitis B vaccine is both safe and effective,[8–10] and universal immunization of infants at birth is the most cost-effective strategy for China.[11] Although immunization at birth is routine in many city hospitals, efforts in rural areas are hampered by the cost of the vaccine and difficulties in transportation, storage, and training.

Our survey focused on village doctors because of the important role they play in the control of iatrogenesis and in the implementation of vaccination and health education programs at the village level. Like the "barefoot doctors" who preceded them, the village doctors are first-line health care providers in rural areas. However, in recent years there has been an effort to upgrade their training, and in many areas where the cooperative medical care system has been dismantled, they work on a fee-for-service basis.[12] Our survey investigated village doctors' sterilization practices, frequency of injection use, current vaccination practices, and knowledge of hepatitis B and counseling skills.

Methods

Setting

Minhou is a rural, middle-income county surrounding Fuzhou, the capital city of Fujian Province. The county is composed of 294 villages divided into 15 townships. Most villages no longer have a fully cooperative medical care system, and simple fee-for-service is common. A 1990 survey by the Minhou Epidemic Prevention Station of 1862 children aged 4 through 6 years found 14.9% to be HBsAg positive (i.e., carriers of the hepatitis B surface antigen). At present there are no data available to establish the carrier rate among adults. The Fujian Epidemic Prevention Station reports the prevalence of HBsAg positivity in the province as a whole to be 14.93%. In Minhou, as in most other rural areas in the province, hepatitis B has not been targeted as a priority health

Serena Clayton is with the School of Public Health, University of California, Los Angeles. Honglian Yang and Jihui Guan are with the Institute for Health Education of Fujian Province, Fuzhou, Fujian, People's Republic of China. Zhongsu Lin and Rui Wang are with the Minhou Epidemic Prevention Station, Minhou, Fujian.

Requests for reprints should be sent to Serena Clayton, MA, Department of Community Health Sciences, School of Public Health, Center for Health Sciences, UCLA, Los Angeles, CA 90024-1772.

This paper was accepted June 22, 1993.

Editor's Note. See related editorial by Sidel (p 1665) in this issue.

TABLE 1—Village Doctors' Administration of Medications by Injection in Seven Hypothetical Cases

	% Doctors Who Would Administer One or More Medications by Injection	% All Medications Used That Would Be Administered by Injection
Middle-aged farmer/thiamine deficiency	77.0	58.7
2-year-old/mumps	76.2	22.2
1-year-old/upper respiratory infection	62.3	25.3
4-year-old/anemia	53.1	24.5
2-year-old/ echovirus diarrhea	43.5	26.0
Married woman/ early pregnancy	42.7	23.6
Conjunctivitis	29.6	12.8

TABLE 2—Total Number of Hypothetical Injections and Total Knowledge Score, by Respondent and Health Station Characteristics

Characteristic	Level (n)	No. Injections[a] Mean	No. Injections[a] Statistics	Knowledge Score[b] Mean	Knowledge Score[b] Statistics
Education	Middle school (153)	4.01	$t = -1.73$	18.51	$t = -2.33$
	High school (107)	4.45	$P = .08$	19.22	$P = .02$
Medical study	≤ 2 y (162)	4.06	$t = -1.22$	18.81	$t = .13$
	> 2 y (98)	4.36	$P = .23$	18.78	$P = .90$
Age	< 38 y (137)	4.14	$t = .46$	18.39	$t = -2.90$
	> 37 y (123)	4.25	$P = .65$	19.25	$P << .01$
Village doctor certificate	No (52)	4.33	$t = .54$	17.60	$t = -4.14$
	Yes (206)	4.16	$P = .59$	19.10	$P << .01$
Years in practice	> 17 (143)	4.06	$t = -1.22$	18.46	$t = -.252$
	> 16 (117)	4.36	$P = .23$	19.21	$P = .01$
First-rate station	No (171)	4.08	$t = .66$	18.61	$t = -1.72$
	Yes (89)	4.25	$P = .51$	19.16	$P = .09$

[a]Total number of injections that would be administered in seven hypothetical cases.
[b]From a total score of 24.

problem. Consequently our findings are probably generalizable to many similar rural areas.

Procedure

Of the 294 villages in Minhou, 260 have village health stations. The head doctor from each health station completed a confidential, self-administered questionnaire that included items on demographics; sterilization, treatment, and counseling practices; and knowledge of HBV transmission, symptoms, and prevalence. The 100% response rate is not unusual in China, where, under the authoritarian system of government, a request for participation is generally viewed as a requirement. One hundred eighty-two doctors completed the questionnaire in one of the 15 township centers during a regular meeting and were reimbursed for transportation and a meal. Seventy-eight doctors completed the questionnaire when a data collector came to their village.

Subjects

Most of the subjects (81.2%) were male and the mean age was 37 years. Most were middle school (56.9%) or high school (38.5%) graduates. The amount of medical training varied: 1 year, 23%; 1 to 2 years, 40%; 2 to 3 years, 19%; more than 3 years, 19%. Eighty percent of the doctors had the "village doctor certificate," which indicates that they have met the national standard for medical knowledge at this level. The mean number of years in practice was 16.2. About one third (34.2%) of the doctors reported that their health station was classified as "first-rate," indicating that it has three rooms, a certified village doctor, a female practitioner, all designated essential drugs, and all national regulations for rural health care.

Results

Sterilization Practices

Most of the doctors (98.1%) reported never using the same needle to inject two patients without sterilizing it. However, only 23.5% reported that they were always able to adhere to the national guideline of "one person, one needle and syringe," probably because syringes are often separated from the needle and reused without sterilization. With respect to methods of sterilization, 59.6% reported sterilizing needles with a pressure cooker and 38.5% reported boiling needles. Eleven doctors (4.2%) reported boiling needles for an insufficient length of time (5–15 minutes). Only one doctor reported using a chemical method of sterilization and four (1.5%) reported washing needles with boiled water. Thus, of 260 health stations, the head doctor at 15 (5.8%) reported an inadequate method of sterilization.

Sterilization of acupuncture needles poses a much greater problem, because they are damaged when brought to high temperatures. National guidelines call for sterilization with glutaraldehyde. Yet of the 76% of doctors who reported that acupuncture is practiced "sometimes" or "often" at their health station, 91.4% sterilized needles with alcohol, which does not reliably destroy HBV. Chemical or other methods of sterilization that do not conform to the national guidelines were reported by 8.6% of the doctors.

Injection Frequency

Subjects were presented with the symptoms and diagnoses for seven hypothetical cases, all of which could be treated with oral medication. The seven cases were upper respiratory infection in a 1-year-old (with fever, congestion, and light cough); mumps in a 2-year-old; nutritional microcytic anemia in a 4-year-old; thiamine deficiency in a middle-aged farmer; early pregnancy in a married woman of unspecified age; echovirus-caused diarrhea in a 2-year-old; and acute conjunctivitis in a patient of unspecified age. The respondents were asked to list the medications and methods of delivery they would use.

For all seven cases together, a mean of 4.19 injections would have been administered (SD = 2.00). Table 1 presents the percentage of respondents who would administer one or more medications by injection and the percentage of all medications used that the respondents would administer by injection. There were no statistically significant bivariate associations between the total number of injections administered for all seven cases and various characteristics of the respondent

and the health station. Table 2 presents the mean number of injections that would be given for the upper and lower (or yes/no) level of each variable.

The doctors were also asked about patients' preferences for methods of drug delivery and the profitability of prescribing oral medications vs giving injections. Most (89.2%) said that patients prefer oral medications. However, when asked what type of drug delivery parents prefer for infants, the reported preferences were oral medications, 28.5%; injections, 21.5%; and both, 49.2%. These responses were grouped into two categories: oral vs injections or both. In three of the four hypothetical conditions dealing with small children, a greater percentage of doctors who felt parents preferred injections or both would administer injections than would doctors who felt parents preferred oral medications: 1-year-old with respiratory infection (68% vs 49%), 4-year-old with anemia (56% vs 47%) and 2-year-old with echovirus-caused diarrhea (46% vs 37%). This difference was statistically significant only in the first case ($\chi^2 = 8.5, P << .01$). With respect to profitability, 91.9% of the subjects reported that it was more profitable for them to administer the same medication by injection than orally.

Vaccination and HBV Testing

Only 11.2% of the doctors reported that most pregnant women in their village are tested for HBV; 60.8% reported that none are tested. The most common reason for women's not receiving the test, given by 58.5% of the doctors, was that the test cannot be performed by the township hospital (58.5%); 34.6% reported that women do not want to be tested, and only 3.1% reported that doctors do not recommend the test.

Vaccination of infants was slightly more common than prenatal testing: 16.2% of the doctors reported that most infants are vaccinated and 48.1% reported that almost none are vaccinated. Of the doctors from first-rate health stations, 64% reported that "most" or "a few" infants are vaccinated, compared with 46% of those from stations that are not first-rate ($\chi^2 = 8.05, P = .02$). The reasons given for nonimmunization were unavailability of the vaccine (68.1%) and cost (24.6%).

Knowledge of HBV

Knowledge of modes of HBV transmission was high. More than 90% knew that HBV could be transmitted by needles, acupuncture, blood products, dental equipment, shared toothbrushes, mouth-to-mouth feeding of infants, and other contacts within the family. Similarly, 90% or more knew that HBV could not be contracted from being out in the sun or catching a chill. Although 87.3% knew that HBV could be sexually transmitted, only 38.5% believed that condoms would prevent transmission. Incorrect responses appeared to be related to confusion with the fecal-oral transmission of hepatitis A. Sixty to sixty-three percent believed that not washing hands after going to the bathroom and eating off dirty dishes could transmit HBV; 36.2% believed that hepatitis B could be contracted by eating spoiled food. Knowledge of the consequences and prevalence of HBV infection was not high: 43.1% did not know that HBV causes liver cancer, and 23.5% believed that hepatitis B is easily cured. Only 24.2% correctly identified the percentage of carriers as 10% to 30%; 41.2% underestimated the rate at 1% to 10%, and 32.3% overestimated the rate at more than 50%.

Table 2 presents the bivariate relationships between total knowledge scores (the sum of 24 true/false items) and various respondent characteristics. Although having more education, having more years in practice, having the village doctor certificate, and being older were all associated with higher knowledge scores, the actual differences in scores were very small. Together these four variables explained only 11.0% of the variation in the total knowledge score in a multiple linear regression model ($F = 7.85, P << .01$), and the village doctor certificate alone accounted for 6.2%.

To determine how knowledge of hepatitis B might be put into practice, we asked the respondents to write down the advice they would give to an HBV carrier. Although our intent was to elicit advice to an asymptomatic carrier, the responses suggest that the question was interpreted to refer to a person with acute HBV infection. Table 3 presents the responses given by 5% or more of the respondents, divided into two categories: those that relate to preventing the transmission of HBV to others and those that relate to improving the health of the carrier/patient. Because of the possible misinterpretation, we restrict our comments to the former subset of responses.

Although the true/false items revealed high knowledge of the modes of transmission of HBV, this knowledge was not employed when formulating advice. The most effective measure, vaccination, was mentioned by only 8.5% of the respondents. Furthermore, only nine doctors (3.5%) would caution a carrier against donating blood. One said that women carriers should not become pregnant, one mentioned use of condoms, and four cautioned against sexual activity. The high percentage of respondents (36.2%) who suggested sterilizing excrement again suggests confusion of hepatitis B with hepatitis A.

TABLE 3—Advice Given to a Hypothetical Hepatitis B Carrier/Patient by 5% or More of Respondents

Advice	%
To prevent transmission	
Separate/do not share toothbrush, eating utensils	84.6
Sterilize, use chlorine bleach for excrement	36.2
Do not go to public places (bathroom, theater, restaurant)	21.9
Vaccinate against hepatitis B virus	8.5
Wash hands after going to the bathroom	5.0
To improve health of carrier/patient	
See doctor/have blood tests/get proper treatment	47.7
Rest/avoid overworking	33.8
Pay attention to nutrition	25.8
Do not eat fatty foods	6.9
Abstain from alcohol or tobacco	6.9
Exercise	5.8

Discussion

Our self-report data indicate widespread inadequacy of sterilization of acupuncture needles and syringes, but not hypodermic needles. However, because acupuncture is rarely practiced on young children, it may not contribute greatly to the rate of chronic HBV infection. Similarly, the contribution of syringe reuse to total HBV carriage is unknown. However, studies demonstrating the resilience of HBV and the facility with which contamination can occur[13,14] suggest that a much more thorough, observational study of medical practices in rural areas is needed. Needles, for example, can easily be contaminated after sterilization.

Our findings of frequent injection use corroborate previous observations in both China[15,16] and other developing countries.[17] The use of injections was not associated with specific characteristics of the doctor or health station; however, doctors were more likely to treat children with injections if they believed the parents

preferred injections. Efforts to reduce the use of injections must focus simultaneously on patients' expectations and providers' practices to address the complex interaction that shapes treatment practices.

Reported levels of hepatitis B vaccination were higher than expected: 16.2% of the village doctors reported that most infants are vaccinated. China is currently in the process of expanding production of the hepatitis B vaccine to increase availability and reduce costs. If these efforts are successful, it is likely that many more parents will be willing to purchase the vaccine for their newborns. Further research into the methods of vaccine delivery and parental knowledge and attitudes is needed for the expansion of immunization coverage.

Doctors at the village level appear to have a good understanding of the transmission of hepatitis B, although there is some confusion with hepatitis A. Not surprisingly, doctors who have not obtained the village doctor certificate appear to have the greatest need for further education about hepatitis B. Otherwise, we did not find any other predictors of practical significance. The lack of useful advice given to a hypothetical carrier/patient suggests that information on HBV has been memorized but not put into practice.

Village doctors are trained through study and practicum, primarily at the township level. Although they participate in regular in-service training, there is little supervision of their activities in their own villages. Our findings suggest that although this system has been reasonably successful with respect to knowledge of hepatitis B, it has not adequately addressed the area of practice. Further training in counseling of HBV carriers and a review of treatment practices should be conducted in conjunction with hepatitis B control efforts. □

Acknowledgment

This research was funded by the National Cancer Institute (CA 49565) Cancer Education Program for Chronic Disease Prevention.

References

1. Xu ZY. Epidemiology of hepatitis B. In: Wen YM, Xu ZY, Melnick JL, eds. *Viral Hepatitis in China: Problems and Control Strategies.* New York, NY: Karger;1992: 99–110. Monographs in Virology, 19.
2. Li Y. Epidemiologic studies of viral hepatitis in China. *Chin J Microbiol Immunol.* 1986;(suppl):1–15.
3. Beasley RP, Hwang LY. Epidemiology of hepatocellular carcinoma. In: Vyas GN, Dienstag JL, Hoofnagle JH, eds. *Viral Hepatitis and Liver Disease. Proceedings of the 1984 International Symposium on Viral Hepatitis.* New York, NY: Grune & Stratton; 1984:209–224.
4. Editorial Committee for the Atlas of Cancer Mortality in the People's Republic of China. *Atlas of Cancer Mortality in the People's Republic of China.* Shanghai, China: China Map Press; 1979.
5. Maynard JE, Kane MA, Hadler SC. Global control of hepatitis B through vaccination: role of hepatitis B vaccine in the expanded programme on immunization. *Rev Infect Dis.* 1989;11(suppl 3):S574–S578.
6. Li RC, Ding ZR, Gong J, et al. Cited in: Xu ZY. Epidemiology of hepatitis B. In: Wen YM, Xu ZY, Melnick JL, eds. *Viral Hepatitis in China: Problems and Control Strategies.* New York, NY: Karger;1992:99–110. Monographs in Virology, 19.
7. Beasley RP, Hwang LY, Lin CC, et al. Incidence of hepatitis B virus infections in preschool children in Taiwan. *J Infect Dis.* 1982;146:198–204.
8. Wong VCW, Ip HMH, Reesink HW, et al. Prevention of the HBsAg carrier state in newborn infants of mothers who are chronic carriers of HBsAg and HBeAg by administration of hepatitis B vaccine and hepatitis immunoglobulin. *Lancet.* 1984;1: 921–926.
9. Xu ZY, Liu CB, Francis DP, et al. Prevention of perinatal acquisition of hepatitis B virus carriage using vaccine: preliminary report of a randomized, double-blind placebo-controlled and comparative trial. *Pediatrics.* 1985;76:713–718.
10. Stevens CE, Taylor PE, Tong MJ, et al. Yeast-recombinant hepatitis B vaccine: efficacy with hepatitis B immune globulin in prevention of perinatal hepatitis B virus transmission. *JAMA.* 1987;257:2612–2616.
11. Xu ZY, Margolis HS. Determinants of hepatitis B vaccine efficacy and implications for vaccination strategies. In: Wen YM, Xu ZY, Melnick JL, eds. *Viral Hepatitis in China: Problems and Control Strategies.* New York, NY: Karger;1992:87–98. Monographs in Virology, 19.
12. Hsiao WC. Transformation of health care in China. *N Engl J Med.* 1984;310:932–936.
13. Lauer JL, VanDrunen NA, Washburn JW, Balfour HH. Transmission of hepatitis B virus in clinical laboratory areas. *J Infect Dis.* 1979;140:513–516.
14. Bond VW, Favero MS, Peterson NJ, Gravelle CR, Ebert JW, Maynard JE. Survival of hepatitis B virus after drying and storage for one week. *Lancet.* 1981;1:550–551.
15. Sun YD. Cited in: Xu ZY. Epidemiology of hepatitis B. In: Wen YM, Xu ZY, Melnick JL, eds. *Viral Hepatitis in China: Problems and Control Strategies.* New York, NY: Karger;1992:99–110. Monographs in Virology, 19.
16. Kleinman A. *Patients and Healers in the Context of Culture.* Berkeley, CA: University of California Press; 1980.
17. Reeler AV. Injections: a fatal attraction? *Soc Sci Med.* 1990;31:1119–1125.

Patient-to-Patient Transmission of Hepatitis B in a Dermatology Practice

ABSTRACT

Objective. The purpose of the study was to investigate infection with hepatitis B virus (HBV) and human immunodeficiency virus (HIV) among patients in a dermatology practice.

Methods. Historical cohort, matched case–control, and cross-sectional survey methods were used.

Results. The age-specific incidence of reported HBV infection in the practice from 1985 through 1991 was more than 12 times the expected rate. The dermatologist was not an HBV carrier. He practiced neither universal precautions nor sterile surgical technique. Seroprevalence of markers for HBV infection was highest (36.8%) among patients who had had surgery on the same day that HBV was apparently acquired by an index case; seroprevalence was near the expected background level for patients not exposed to index cases. Of HBV-infected patients with known dates of onset, 72% had had surgery during their incubation periods. All of 30 HBV antigen specimens tested were of the same subtype. None of the patients tested, including 74 patients exposed to surgery on the same day as a patient with acquired immunodeficiency syndrome, had evidence of HIV infection.

Conclusions. HBV, but not HIV, was transmitted from patient to patient by the dermatologist's failure to apply either universal precautions or sterile surgical technique. (*Am J Public Health.* 1993;83:1689–1693)

W. Gary Hlady, MD, MS, Richard S. Hopkins, MD, MSPH, Terry E. Ogilby, RN, BSN, and Sandra T. Allen, BSN

Introduction

Concerns regarding the risk of acquiring hepatitis B virus (HBV) and other blood-borne infections during medical and dental procedures have increased owing to the epidemic of acquired immune deficiency syndrome (AIDS) and apparent transmission of the human immunodeficiency virus (HIV) in a dental practice.[1] Transmission of HBV in health care settings is much more common than transmission of HIV. Outbreaks of HBV infection have resulted from transmission of the virus from infected health care workers to their patients,[2–4] and patient-to-patient transmission has been described in nosocomial outbreaks[5–7] and in alternative-care settings, such as acupuncture clinics.[8–10] In this report, we describe a large outbreak of HBV infection that was the first identified as resulting from patient-to-patient transmission in a conventional outpatient medical practice.

In September 1991, public health officials in Lee County, Florida, recognized that eight patients with acute HBV infection reported since 1985 had visited a Fort Myers dermatologist prior to onset of their disease. The dermatologist had no history of acute hepatitis or hepatitis B immunization. In 1987, he was found to be hepatitis B surface antigen (HBsAg) negative and hepatitis B surface antibody (HBsAb) positive, indicating that he was not a carrier but had been previously exposed to HBV. Patient records indicated that he had seen more than 10 000 patients since 1985, most of whom were White and older than 54 years of age.

Although the investigation was initiated by reports of HBV-infected patients, there was concern that HIV may have been transmitted as well.

Methods

Practice Evaluation

The dermatologist was interviewed regarding office practices and procedures, and he was instructed regarding universal precautions to control the spread of HBV and other blood-borne disease.[11] He was subsequently observed with 16 consecutive patients by one of the authors (W. G. H.) to determine whether universal precautions were being followed. Office autoclaves were tested for proper timing, temperature, and pressure.

Incidence Study

Only records for active patients and patients whose charts had been open in 1985 or later were available for review. Chart review identified all of the dermatologist's patients who had been reported to the Lee County Public Health Unit with acute HBV infection from January 1985 through September 1991, including those for whom exposure to the dermatologist was not mentioned in the original case reports. The findings were used to calculate the proportion of all reported cases of acute HBV infection who were patients of

W. Gary Hlady and Richard S. Hopkins are with the Epidemiology Program, State Health Office, Florida Department of Health and Rehabilitative Services, Tallahassee, Fla. Terry E. Ogilby and Sandra T. Allen are with the Epidemiology Office, Lee County Public Health Unit, Fort Myers, Fla.

Requests for reprints should be sent to W. Gary Hlady, MD, MS, Epidemiology Program, State Health Office, Florida Department of Health and Rehabilitative Services, 1317 Winewood Blvd, Tallahassee, FL 32399-0700.

This paper was accepted May 18, 1993.

Editor's Note. See related editorial by Mosley (p 1664) in this issue.

the dermatologist and to estimate the incidence of reported cases in his practice.

Case–Control Study

Early in the course of the investigation, we conducted a case–control study to test the hypothesis that HBV-infected patients were more likely than other patients to have had a surgical procedure during the incubation period of the disease (45 to 180 days before onset of symptoms). Case patients were persons reported with acute HBV infection with a known date of onset for whom treatment records were available from the dermatologist. Three randomly selected control patients with available treatment records were matched with each of 20 case patients by age (± 10 years) and year of visit—that is, controls visited the dermatologist at least once during the same year that the corresponding cases acquired their infections. The control patients were randomly selected from among patients of the dermatologist who were not reported with hepatitis B.

Relative risk was estimated by the matched odds ratio (OR) with a 95% confidence interval (CI).[12] The probability (*P*) that the observed matched odds ratio might have occurred by chance was obtained by stratified Mantel-Haenszel analysis, with each stratum consisting of one case and three matched controls.[13]

Cross-Sectional Study of HBV Seroprevalence

On the basis of the results of the case–control study, we defined patients at high risk of exposure as those who had had surgery within 1 day of surgery on HBV-infected patients during their incubation periods and those who had had surgery on the day of or the day after surgery on HBV-infected patients subsequent to their incubation periods. Nearly all of the surgical procedures consisted of excisional biopsies, but no appointment records were available for years prior to 1991; therefore, most of the high-risk patients were identified through pathology reports, which indicated the date on which a surgical specimen was submitted. High-risk patients were notified individually by mail of the possibility that they had been exposed to HBV and "other blood-borne disease" and were urged to appear for testing. In addition, media publicity attracted a large number of patients for testing who had not been identified as high risk. All patients who appeared for testing were offered tests for HBsAb, HBsAg, and antibodies to HIV.

Testing was provided free of charge from November 7 to December 18, 1991. Sera were screened for HBsAg and HBsAb with standard enzyme immunoassay reagents (Abbott Laboratories). Positive HBsAg results were confirmed by virus neutralization. In addition, the HBsAg in specimens from 30 patients—selected because they had known dates of onset or documented exposures to other infected patients—were subtyped, using monoclonal antibodies with restricted specificity in an enzyme immunoassay.[14] Screening for antibodies to HIV was also performed by enzyme immunoassay (Genetic Systems).

Informed consent was obtained for all testing, and pre- and posttest counseling was given to all patients who consented to be tested for HIV. All patients with positive test results were interviewed regarding hepatitis B vaccination, compatible clinical illness, onset dates, and possible risk factors, such as blood transfusions, injection drug use, male homosexuality, and contact with an infected family member or sexual partner.

The test results from patients without a history of hepatitis B vaccination were used to calculate prevalence of serum markers for infection for several categories of exposure. Only patients with a single exposure in a single given category were included in these calculations, with one exception—all unvaccinated patients not identified as being at high risk of infection were included in a comparison group, regardless of the number of times they were seen in the practice. Standard methods based on the binomial probability distribution were used to calculate confidence intervals for prevalence and to test the significance of differences in prevalence ratios between categories.[15]

A number of patients were tested by private health care professionals. Because not all results were reported to us, patients tested privately who did report their positive results to us were included in case counts but were not included in the calculation of prevalence ratios.

Cross-Sectional Study of HIV Seroprevalence

Local AIDS surveillance staff searched patient files in the dermatology practice to find records for patients who had been reported with AIDS and to abstract their treatment histories. Patients who had had surgery on the day of or the day after surgery on a patient with AIDS were identified as being at high risk of exposure to HIV. These patients were then notified and tested in exactly the same manner described above.

Results

Reported Incidence

A total of 30 of the dermatologist's patients were reported with acute HBV infection from 1985 through 1991. Twenty-three were older than 54 years at the time of diagnosis, representing 52% (23/44) of all HBV-infected patients reported from Lee County in that age group. The estimated incidence of reported HBV-infected patients older than 54 years in the dermatology practice was 71.8 per 100 000 person-years, more than 12 times the rate of 5.9 observed for the same age group in the county as a whole during the same period.

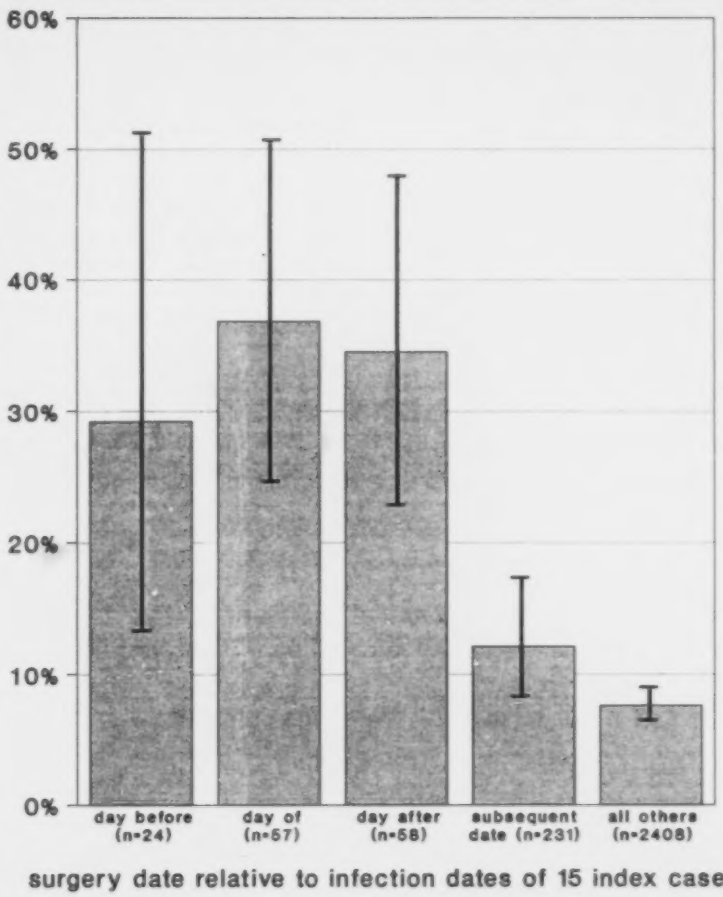
Case–Control Study

Eighty percent (16/20) of the case patients and 40% (24/60) of the matched control patients had had surgery during the case patients' incubation periods (matched OR = 9.00; 95% CI = 1.82, 44.45; $P = .003$). Subsequently, six of the control patients appeared for screening; two were identified as having been infected with HBV. Neither had a known date of onset, and both had had surgery in the practice during the incubation periods of their matched case patients. When these two control patients were deleted from the analysis, the matched OR was 9.33 (95% CI = 1.93, 45.04; $P = .002$).

Cross-Sectional Study of HBV Seroprevalence

Surgery dates of 22 HBV-infected patients (index case patients) were used to identify a total of 736 patients as being at high risk of infection. Four hundred (54%) of these high-risk patients, excluding those who had received hepatitis B vaccine, were tested for HBsAb and HBsAg. Twenty percent (80) tested positive by one test or the other, compared with 7.6% (183) of 2408 other unvaccinated patients who appeared for testing as a result of media publicity or because they were identified as high risk only for infection with HIV (risk ratio = 2.63; 95% CI = 2.07, 3.35; $P << .01$).

The seroprevalence results were further divided into five categories of exposure, shown in Figure 1. Seroprevalence of markers of HBV infection was highest, at 36.8% (95% CI = 24.7, 50.7), among those patients who had had surgery on the same day that an index case patient ap-



Note. Infection dates of index cases are dates of surgery during the case patients' incubation periods. Except for the "all others" category, only patients with a single surgical exposure in a single given category were included in the prevalence calculations. Some patients in the "all others" category may have been exposed to HBV-infected patients in the practice who had not been identified at the time of screening.

FIGURE 1—Prevalence of serum markers for HBV in five risk groups.

parently acquired his or her infection. Seroprevalence was also relatively high among those who had had surgery the day before or the day after an index case patient's apparent infection date. Having a surgical procedure on days when index case patients had subsequent surgical procedures was also associated with a prevalence significantly greater than that observed for patients not in a high risk category (12.1% vs 7.6%, $\chi^2 = 5.86$, $P = .02$).

Exposure Histories

A total of 305 unvaccinated patients in the dermatology practice, including index case patients, had serological evidence of infection with HBV. Of these, 69 had known dates of onset and patient records available for review. These patients had an average age at onset of 67 years (range: 29–91); 38 were men and 31 were women. Patient records documented surgery by the dermatologist during the incubation periods for 50 (72%) of the HBV-infected patients with known dates of onset, including 2 who eventually died of the disease. The mean incubation period was 126 days (range: 58–168 days).

Of the 19 HBV-infected patients with known dates on onset and available treatment records who had no record of surgery during their incubation periods, 2 were spouses of patients with HBV infection who had had surgery in the practice during their incubation periods, 1 was a homosexual man, and 3 had been seen during their incubation periods but there was no record of a surgical procedure. The remaining 13 patients had no known risk factors for HBV infection.

None of the HBV-infected patients with documented surgery in the practice during their incubation periods had any other risk factors for HBV infection.

HBsAg Subtyping

The epidemic curve (Figure 2) shows 70 of the 72 HBV-infected patients with known dates of onset. The 2 patients not shown were both chronic carriers of HBV, one with onset in 1945 and the other with onset in 1965. Subtyping of HBsAg from the patient with onset in 1945, along with 29 others, showed that all were subtype *adw2*.

HIV Exposures and Seroprevalence

The practice included 22 patients with AIDS; all had died by the time of the investigation. Two had no identified risk factors. Each of these 2 had had only one surgery in the practice, at 5 and 15 months prior to their deaths. A total of 11 surgical procedures had been performed on 9 of the patients with AIDS in the practice, but none on a day that HBV was apparently transmitted. One hundred eighty-nine patients had had surgery on the same day as surgery on an AIDS patient or on the day after, and 74 of these, including 2 who were exposed twice, were tested for antibodies to HIV. Results for all of the 74 exposed patients and all of the other 2621 patients tested were negative.

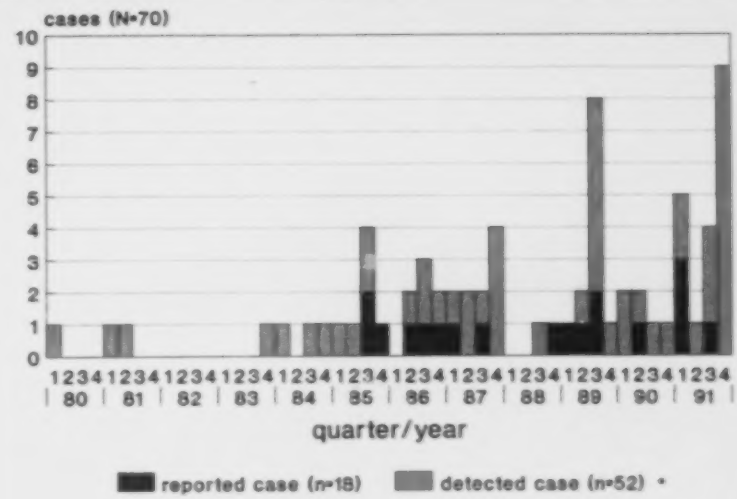
Possible Vehicles for HBV Transmission

Possible vehicles for HBV transmission between patients included the dermatologist's hands, injectable solutions in multiple-use vials, electrocautery tips, and liquid nitrogen.

The dermatologist's hands were often contaminated with blood, as he routinely operated without gloves and failed to wash his hands after every patient.

Injectable anesthetic solution (1% lidocaine without epinephrine) in multiple-use 50-cc vials may have been contaminated. Only one 5-cc syringe was used per patient. If additional solution was required, the same syringe was refilled through a needle that remained in the vial and was common to all other patients seen that day.

Electrocautery tips were neither changed nor cleaned between patients, and cotton-tipped liquid nitrogen applicators were occasionally returned to the res-



* Patients tested November 7 - December 18, 1991.

Note. Reported cases were investigated by the county public health unit at the time of diagnosis. Detected cases were identified by serologic screening for HBsAb and HBsAg, with onset dates determined by patient history of jaundice or acute illness diagnosed as hepatitis B.

FIGURE 2—Onset dates for 70 HBV-infected patients in a dermatology practice, 1980 through 1991.

ervoir after use (although liquid nitrogen was not used in the surgical procedures examined by our investigation). We found no other problems with office sterilization equipment or procedures for cleaning and sterilizing instruments.

Discussion

Investigation of this outbreak was complicated by several factors. The infections occurred over a long period of time among a large number of people. Many infections were asymptomatic, and reporting of symptomatic infections was incomplete. Nevertheless, a strong argument can be made that patient-to-patient transmission of HBV occurred with disturbing frequency. The observation that HIV was not also transmitted in this setting is consistent with earlier estimates of the relative infectivity of the two viruses[16–18] and further emphasizes the importance of HBV as a blood-borne pathogen in the health care setting.

The incidence calculations confirmed the existence of an outbreak. The number of person-years of exposure to the practice was necessarily a rough estimate for calculation of a rate of infection. However, any error introduced was unlikely to account for the 12-fold difference between the estimated rate of reported HBV infection in the practice and that observed for the county as a whole.

The hypothesis that infection was transmitted via surgical procedures was based on biologic plausibility and the dermatologist's description of his surgical technique. It is possible that some control patients in the case–control study, in addition to the two we identified, may have been infected with HBV, but such misclassification would only have reduced our estimate of the matched odds ratio. The level of risk indicated by the matched odds ratio, therefore, represents a minimum estimate.

The results of the case–control study suggested high-risk subsets of patients to actively screen for evidence of infection, both with HBV and with HIV, and the validity of this classification was confirmed by the test results. Seroprevalence of markers for HBV infection was significantly elevated among high-risk patients in a pattern predicted by our hypothesis that patient-to-patient transmission occurred during surgical procedures.

The elevated prevalence of markers of HBV among patients who had had surgery the day before or the day after apparent infection of an index case patient is consistent with transmission by a vehicle, such as contaminated solutions for injection, that could have been used on successive days. Multiple-use vials for injection were identified as the likely vehicle of infection with HBV in a previous outbreak, and the potential certainly existed for contamination of the 50-cc vials of local anesthetic solution as they were used in this dermatology practice.[6]

Declining seroprevalence among patients exposed to days of subsequent surgery on index case patients probably reflects waning antigenemia. And seroprevalence in the low-risk patients with no known exposure to an index case patient (7.6%) was close to the 7% found in Whites older than 54 years in a representative sample of the general US population.[19]

The time of introduction of HBV into the practice is unknown. Two carriers of HBV with long histories of infection were identified in the practice, and the dermatologist himself may have been a source at some time prior to 1987. However, there is good evidence that HBV transmission continued from patient to patient. The logarithmically expanding epidemic curve is characteristic of patient-to-patient transmission, although this pattern may have been exaggerated by detection and recall biases. There were several potential vehicles for patient-to-patient transmission. A high proportion of HBV-infected patients were temporally linked to the practice and to other infected patients, and all of the specimens tested for antigenic subtype were identical. The *adw2* HBV subtype accounts for approximately 88% of the HBV infections in the southeastern United States, making the probability of finding *adw2* by chance alone in all 30 specimens from this outbreak equal to 1 in 50 ($P = .02$).[20]

In summary, we found more than 12 times the expected rate of patients reported with acute hepatitis B in a dermatology practice; most cases had had surgery in the practice during their incubation periods. Transmission from patient to patient appeared to result from failure to observe either universal precautions or sterile surgical technique. Finally, although HBV appears to have been freely transmitted, we found no evidence of transmission of HIV.

These findings strongly support efforts to increase awareness of blood-borne infections among physicians and to ensure that universal precautions are universally practiced. They also emphasize the importance of reporting blood-borne diseases to public health authorities, along with detailed exposure histories, and demonstrate once again that continuing review

of case reports by health departments is useful in identifying and controlling outbreaks. The outbreak described here was so insidious and of such long duration that it could easily have escaped detection if only the numbers of reported cases had been examined, without an epidemiologic review of potential sources of infection. Finally, consideration should be given to establishing standards of care for office-based procedures that would prevent the spread of blood-borne disease in much the same way that standards for food handling are used to prevent the spread of enteric disease. □

Acknowledgments

We are grateful to all those who assisted with this investigation, especially E. E. Buff, MS; Gail Counts, RN, BS; John J. Naponick, MD, MPH; Elaine Levidow, BS; Irma M. See, MPH; Craig N. Shapiro, MD; Paul D. Swenson, Ph.D; and Bradley A. Woodruff, MD, MPH.

References

1. Ciesielski C, Marianos D, Ou C, et al. Transmission of human immunodeficiency virus in a dental practice. *Ann Intern Med.* 1992;116:798–805.
2. Lettau LA, Smith JD, Williams D, Lundquist WO, Cruz F, Sikes RK. Transmission of hepatitis B with resultant restriction of surgical practice. *JAMA.* 1986;255:934–937.
3. Kane MA, Lettau LA. Transmission of HBV from dental personnel to patients. *J Am Dent Assoc.* 1985;110:634–636.
4. Centers for Disease Control. Outbreak of hepatitis B associated with an oral surgeon—New Hampshire. *MMWR.* 1987; 36:132–133.
5. Polish LB, Shapiro CN, Bauer F, et al. Nosocomial transmission of hepatitis B associated with the use of a spring-loaded finger-stick device. *N Engl J Med.* 1992;326: 721–725.
6. Alter MJ, Ahtone J, Maynard JE. Hepatitis B virus transmission associated with a multiple-dose vial in a hemodialysis unit. *Ann Intern Med.* 1983;99:330–333.
7. Douvin C, Simon D, Zinelabidine H, Wirquin V, Perlemuter L, Dhumeaux D. An outbreak of hepatitis B in an endocrinology unit traced to a capillary-blood-sampling device. *N Engl J Med.* 1990;322: 57.
8. Kent GP, Brondum J, Keenlyside RA, La-Fazia LM, Scott HD. A large outbreak of acupuncture-associated hepatitis B. *Am J Epidemiol.* 1988;127:591–598.
9. Stryker WS, Gunn RA, Francis DP. Outbreak of hepatitis B associated with acupuncture. *Am J Fam Pract.* 1986;22:155–158.
10. Canter J, Mackey K, Good LS, et al. An outbreak of hepatitis B associated with jet injections in a weight reduction clinic. *Arch Intern Med.* 1990;150:1923–1927.
11. Centers for Disease Control. Recommendations for preventing transmission of human immunodeficiency virus and hepatitis B virus to patients during exposure-prone invasive procedures. *MMWR.* 1991; 40(RR-8):1–8.
12. Robins J, Greenland S, Breslow NE. A general estimator of the variance of the Mantel-Haenszel odds ratio. *Am J Epidemiol.* 1986;124:719–723.
13. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22:719–748.
14. Swenson PD, Riess JT, Krueger LE. Determination of HBsAg subtypes in different high risk populations using monoclonal antibodies. *J Virol Methods.* 1991;33:27–38.
15. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley & Sons; 1981;19–32.
16. Henderson DK, Fahey BJ, Willy M, et al. Risk for occupational transmission of human immunodeficiency virus type 1 (HIV-1) associated with clinical exposures: a prospective evaluation. *Ann Int Med.* 1990;113:740–746.
17. Marcus R, CDC Cooperative Needlestick Study Group. Surveillance of health-care workers exposed to blood from patients infected with the human immunodeficiency virus. *N Engl J Med.* 1988;319:1118–1123.
18. Gerberding JL, Bryant-LeBlanc CE, Nelson K, et al. Risk of transmitting the human immunodeficiency virus, cytomegalovirus, and hepatitis B virus to health-care workers exposed to patients with AIDS and AIDS-related conditions. *J Infect Dis.* 1987;156: 1–8.
19. McQuillan GM, Townsend TR, Fields HA, Carroll M, Leahy M, Polk BF. The seroepidemiology of hepatitis B virus in the United States, 1976–1980. *Am J Med.* 1989; 87(suppl 3A):5–10.
20. Dodd RY, Holland PV, Ni LY, Smith HM, Greenwalt TJ. Hepatitis B antigen: regional variation in incidence and subtype ratio in the American Red Cross donor population. *Am J Epidemiol.* 1973;97:111–115.

Salmonella Enteritidis Infections in France and the United States: Characterization by a Deterministic Model

Laurence Watier, PhD, Sylvia Richardson, PhD, and Bruno Hubert, MD

ABSTRACT

Objectives. Epidemiological surveillance can be used to identify problems, suggest hypotheses, and assess the effectiveness of preventive actions. These objectives are pursued in the analysis of the dynamic evolution of *Salmonella enteritidis* infections. In this analysis, development of the epidemic in France and the United States is compared.

Methods. This analysis is based on the formulation of a simple deterministic compartmental model, in which human contact with contaminated food is expressed in terms of a baseline transmission rate.

Results. In France, the baseline transmission rate, stable up until 1986, is multiplied by 2.3 in 1987 and by 4.1 in 1988. There is no evidence of a slowdown of the linear rate of increase. In the Middle Atlantic region of the United States, the linear increase of the epidemic is similar to that observed in France. From 1990 a potential effect of the preventive measures is observed.

Conclusions. The pattern of increase of the baseline transmission rate is similar in France and in the Middle Atlantic region of the United States. However, preventive measures, used in both areas, appeared more effective in the United States. (*Am J Public Health.* 1993;83:1694–1700)

Introduction

Salmonella enteritidis infections are a new and increasing public health problem in most European countries and in North America.[1] Consumption of raw or inadequately cooked eggs has been found to be the major source of infection.[2] Some countries, such as the United States, the United Kingdom, and France, have recently implemented control measures to identify and slaughter flocks of infected laying birds. To evaluate the efficacy of these measures, two sources of information are available: notification of outbreaks and identification of *Salmonella* isolates by a national reference center.

We propose a simple deterministic compartmental model to analyze the dynamic evolution of the total number of *S. enteritidis* isolates in a population, and we estimate this model on data from reference centers in France and the United States. In this model, we distinguish only two groups of individuals, susceptible and infected, since *S. enteritidis* is not known to confer any immunity.[3] A time-dependent transition rate from susceptible to infected is modeled at first in terms of both a baseline transmission rate (β_0) from the animal reservoir to humans and a seasonal multiplicative factor.

Such a model can help to achieve several public health objectives. Firstly, the estimation of β_0 for different time periods enables us to quantify the increase in cases of *S. enteritidis* observed in recent years and to compare the epidemic's development between countries. We show comparative results between France and the Middle Atlantic region of the United States.

Secondly, time components in the modeling of the transition rate from susceptible to infected can also be included to test if the time structure of potential environmental sources of contamination is reflected in the evolution of the indicator. Investigations in France from 1989 to 1990 showed that 52% of *S. enteritidis* outbreaks associated with eggs were due to free-range eggs from small holdings that were noncommercially distributed.[4] Thus, we included in the transition rate an indicator of the seasonal time pattern of egg laying of free-range hens.

Finally, model predictions were calculated to measure the impact of preventive measures taken in France since May 1989 and in the United States since February 1990.[5]

Material and Methods

Data

France. The data used in this study come from the National *Salmonella* Reference Center at the Pasteur Institute. This center receives *Salmonella* isolates from private or public laboratories for serotyping. Monthly notifications of identified *S. enteritidis* isolates are available dating from January 1978. The period analyzed includes 156 months of observations from January 1978 to December 1990. In addition, quarterly data on the number of isolates identified in laboratories other than the National Reference Center are also available and are used to check the stability of the distribution of isolates between the center and other laboratories.

Laurence Watier and Sylvia Richardson are with the Institut National de la Santé et de la Recherche Médicale, Villejuif, France. Bruno Hubert is with the Réseau National de Santé Publique, Saint Maurice, France.

Requests for reprints should be sent to Laurence Watier, PhD, Inserm Unité 170, 16, av Paul Vaillant Couturier, 94807 Villejuif, Cedex, France.

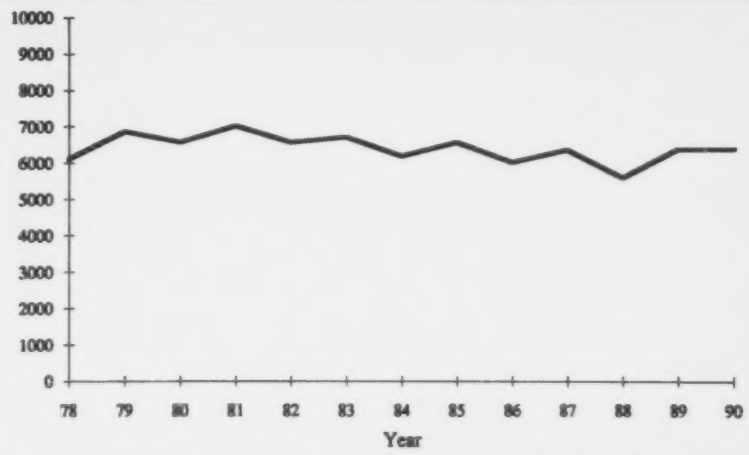
This paper was accepted June 29, 1993.

Editor's Note. See related annotation by Halloran (p 1667) in this issue.

Up until the end of 1988, the number of *Salmonella* isolates other than *S. enteritidis* and *S. typhimurium* was fairly constant (Figure 1); consequently, the observed increase in the number of *S. enteritidis* isolates in 1987 and 1988 (Figure 2) probably corresponds to a real increase in the incidence rather than to improved reporting. The mean-range plot confirms that the years 1987 and 1988 stand out (Figure 3). However, because of media coverage and some reorganization of laboratory analysis practices, two sources of bias were introduced. First, higher awareness of the dangers of *Salmonella* infections is reflected in a 16% increase in the yearly total of *Salmonella* isolates of all serotypes other than *S. enteritidis* and *S. typhimurium*. Second, it is known that in 1989, laboratories outside the National Reference Center started to equip themselves with antisera to be able to identify *S. enteritidis*. Indeed, for each year from 1978 to 1988, the center identified approximately 78% of all reported *S. enteritidis* isolates. The situation changed in 1989, however, and this proportion decreased to 74% and 47% for 1989 and 1990, respectively. Hence, model estimation is based on reference center data only up until the end of 1988. Yearly predictions for 1989 and 1990 are compared with observed yearly numbers after allowing for some bias corrections.

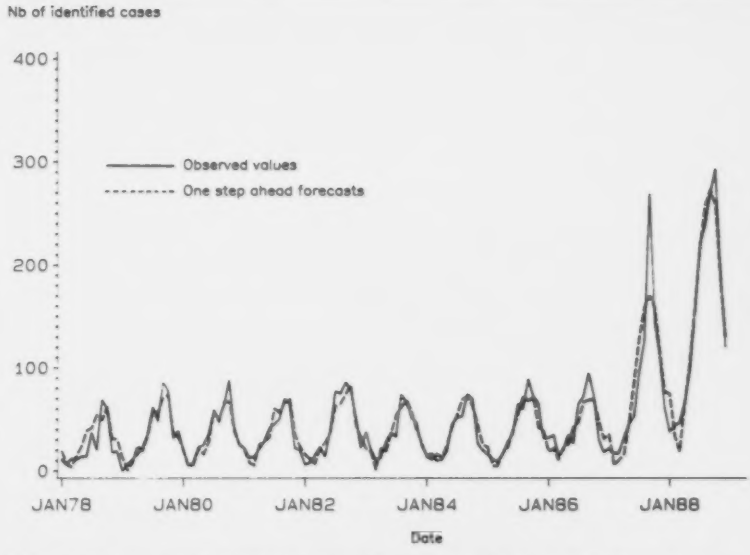
United States. The data analyzed come from the Centers for Disease Control and Prevention in Atlanta and are available for the period 1968 to 1990. The data are based on a passive laboratory-based *Salmonella* surveillance activity. The number of cases reported was found to represent from 1% to 5% of the real incidence of this infection.[6] The Middle Atlantic region of the United States was selected for analysis to maximize comparability with France data, given that similar climatic conditions prevail in both regions. The increase of reported cases started in 1984, providing a long stable period (Figures 3 and 4).

As in France, the system has inherent biases. However, Hargrett-Bean et al.[7] judged the data to be of sufficiently high quality to allow some insight as to the effectiveness of public health intervention.



[a]The values for 1989 and 1990 are obtained after allowing for a 16% mediatic increase.

FIGURE 1—Yearly number of isolates of *Salmonella* in France[a]: all serotypes except *S. typhimurium* and *S. enteritidis*.



[a]The values corresponding to July, August, September, and October 1984 have been corrected because they mistakenly included some outbreak cases.
[b]Model includes a baseline transmission rate (β_0), a seasonal factor (β_1), and a seasonal laying effect (β_2).

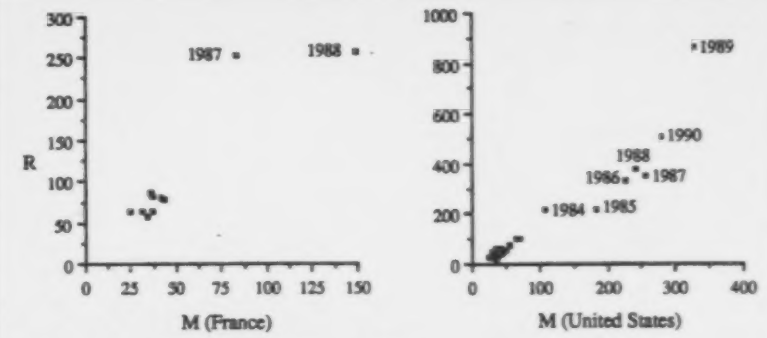
FIGURE 2—Comparison of the observed values[a] and fitted values of model 2[b] for *S. enteritidis* in France from 1978 to 1988.

The Compartmental Susceptible-Infected Model

Person-to-person contact is considered of little importance in *Salmonella* transmission in contrast to classical infectious disease models.[8] Most new cases are related to the animal reservoir through foodstuff contamination. The dynamic evolution of *Salmonella* infections is analyzed using a simple deterministic compartmental model. The population is divided into subgroups, the susceptibles and the infecteds, and the flows between these two subgroups per unit of time are expressed. The model can be represented schematically as follows:

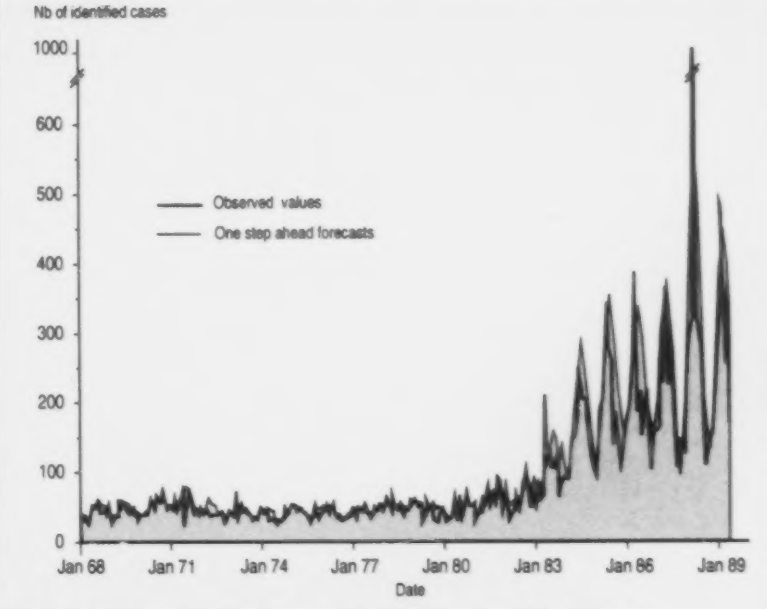
$$\boxed{NS(t)} \xrightarrow{f(t)} \boxed{NI(t)}, \qquad \boxed{NI(t)} \xrightarrow{\gamma} \boxed{NS(t)}$$

where $S(t)$ denotes the proportion of individuals not infected but susceptible to contract *S. enteritidis* infection at time t;



[a]The plot of the range against the mean for each seasonal period.

FIGURE 3—Mean-range plot[a]: mean and range for yearly *S. enteritidis* in France from 1978 to 1988 and in the Middle Atlantic region of the United States from 1968 to 1990.



[a]Model includes a baseline transmission rate (β_0) and a seasonal factor (β_1).

FIGURE 4—Comparison of the observed values and the one-step-ahead forecasts from the fitting of model 1[a] for *S. enteritidis* in the Middle Atlantic region of the United States from 1968 to 1990.

$I(t)$ denotes the proportion of individuals infected with *S. enteritidis* at time t;
N is the total population size, assumed to be constant; and
$f(t)$ and γ are the flows per unit of time between the two subgroups.

Note that $NI(t)$ represents the number of cases of the disease at time t.

The following hypotheses are made throughout:

1. Susceptible individuals become infected per unit of time at a rate $f(t)$, which represents a contact between individuals and contaminated foods; and
2. Infected individuals recover and become susceptible at a constant rate γ, with $1/\gamma$ representing the mean infection time.

Assuming a constant population size, these hypotheses lead to linear differential equations for $S(t)$ or $I(t)$, which can be easily solved (see Appendix).

Two models are proposed for the transmission rate $f(t)$. *Salmonella* multiplication is facilitated by a temperature of 20°C or higher, which explains the increased incidence of infection during summer months. Thus,

$$f(t) = \beta_0\left(1 + \beta_1 \sin\frac{2\pi t}{12}\right) \quad \text{(Model 1)},$$

where β_0 is a baseline transmission rate from the animal reservoir to humans and β_1 is a multiplicative seasonal effect of period 12 (Figure 5).

Investigations in France have tended to incriminate noncommercially distributed free-range eggs from small holdings, even though this category of eggs accounts for only 20% of French production.[4] Climatic conditions between November and April lead to an interruption of laying in nonindustrial free-range hens. This suggests a second model, including a further time component in the transmission rate $f(t)$:

$$f(t) = \beta_0\left(1 + \beta_1 \sin\frac{2\pi t}{12}\right) + \beta_2 g(t) \quad \text{(Model 2)},$$

where $g(t)$ is a periodic indicator function of the laying period and is equal to 1 between April and October, inclusively. The parameter β_2 is referred to as a seasonal laying effect (Figure 6).

Estimation

To be able to estimate $N\beta_0$, β_1, and, in some cases, $N\beta_2$ for fitting model 1 or 2 to the observed values of the number of new isolates $O(t)$ over a period $[t_0, t_1]$, it is necessary to suppose that $N\beta_0$, β_1, and $N\beta_2$ are constant over $[t_0, t_1]$. Clearly this is a reasonable assumption in a stable situation. In a situation in which an epidemic is developing, the time intervals chosen will be relatively short (say a year) so that this assumption holds approximately.

In the fitting of model 1, two coefficients—$N\beta_0$ and NC_1—are estimated with $C_1 = \beta_0\beta_1$ (see Appendix). To take into account possible time autocorrelation of the regression residuals, the coefficients $N\beta_0$ and NC_1 are estimated using a regression model with autocorrelated errors (procedure Autoreg of SAS/ETS[9]) and the value of $\hat{\beta}_1$ is deduced. To obtain the variance of $\hat{\beta}_1$, which is a multiple of $\hat{C}_1/\hat{\beta}_0$, the delta method is used. The fitting of model 2 proceeds along similar lines.

Prediction

In France, the stationary structure of *S. enteritidis* infections, which prevails until 1987, gives way to substantial increases in the following years (Figure 2). The US situation is similar, but the increase starts in 1984 (Figure 4). Under these conditions, forecasting is possible only if one is prepared to model the increase; with our approach, this entails predicting values of $N\beta_0$, β_1, and $N\beta_2$ by prolonging their observed trend.

Predictions are used as benchmarks to assess the impact of preventive measures. The observed values for the years after the start of the preventive measures are compared with the forecasts.

Results

Quantification of the Increase in the Baseline Transmission Rate in France

To confirm that the structure of the data was indeed perturbated in 1987, forecasts for that year were obtained, based on the previous observations. From March 1987 onward, the observed number of isolates lies well above the one-sided 95% upper confidence interval of the forecasts built from the 1978 to 1986 period, indicating that a nonstationary pattern of increase started in March 1987. Model 1 was thus estimated separately for the stable period of 1978 to 1986 and then year by year to characterize the increase.

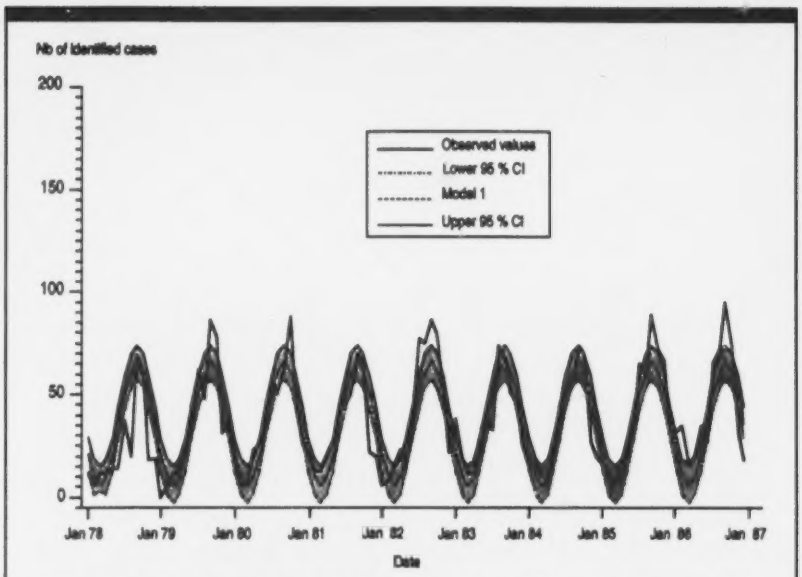
For the stable period, it was necessary to introduce autoregressive parameters at lags 1, 3, 8, 12, and 24. The autocorrelations at lags 1 and 3 reflect an expected short-term correlation, while those at lags 8, 12, and 24 are related to residual seasonality not accounted for by the deterministic part of the model.

All model parameters are clearly significant (Table 1). The seasonal multiplicative effect is fairly constant. On the other hand, there is a substantial increase of the baseline parameter β_0, which is first multiplied by 2.3 in 1987 and then by 4.1 in 1988. The sinusoidal shape of $Nf(t)$ in model 1 is illustrated in Figure 5 for the stable period, with parameters chosen in accordance with Table 1.

Quantification of the Increase in the Baseline Transmission Rate in the US Middle Atlantic Region

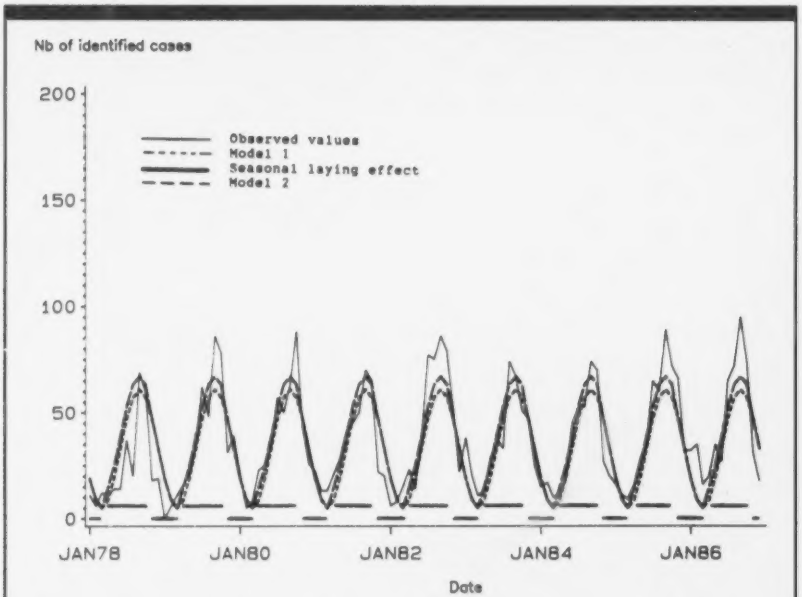
The same compartmental model was estimated separately for the stable period 1968 to 1983 and then year by year. For the stable period, the residuals were modeled with significant autoregressive parameters at lags 1, 3, and 11.

The parameter $N\beta_0$ is clearly significant over the eight periods (Table 2). The increase of β_0 seems to have diminished in 1990. For the stable period, the seasonality is less marked than it was for the French data, but the seasonal parameter β_1 is nonetheless significant. From 1984 onward, the seasonal effect is strengthened.



*$N\beta_0$ and β_1 have been replaced by their estimated values (36.1 and −0.8, respectively).

FIGURE 5—*S. enteritidis* in France from 1978 to 1986: time plot of $Nf(t)$, where $f(t)$ is the transmission rate corresponding to model 1 (together with its 95% confidence limits).*



*$N\beta_0$, β_1, and $N\beta_2$ have been replaced by their estimated values (32.9, −0.9, and 6.1, respectively).

FIGURE 6—*S. enteritidis* in France from 1978 to 1986: time plot of $Nf(t)$, where $f(t)$ is the transmission rate corresponding to model 2 (together with its two components).*

TABLE 1—Estimation of the Parameters for Model 1[a] for *S. enteritidis* in France from 1978 to 1988

Period	$N\hat{\beta}_0$ ($\hat{\sigma}_0$)	R_0[b]	$\hat{\beta}_1$ ($\hat{\sigma}_1$)
1978/86	36.1		−0.8
	(4.1)	1	(0.1)
1987	83.2		−1.0
	(11.3)	2.3	(0.2)
1988	148.7		−0.8
	(7.4)	4.1	(0.1)

Note. Parameter estimates are indicated by a "^" with standard errors in brackets. All parameters were significant at level $P < .001$.
[a]Model includes a baseline transmission rate (β_0) and a seasonal factor (β_1).
[b]R_0 corresponds to the ratio of $N\hat{\beta}_0$ estimated in 1987 or in 1988 to its value for the stable period.

TABLE 2—Estimation of the Parameters for Model 1[a] for *S. enteritidis* in the Middle Atlantic Region of the United States from 1988 to 1990

Period	$N\hat{\beta}_0$ ($\hat{\sigma}_0$)	R_0[b]	$\hat{\beta}_1$ ($\hat{\sigma}_1$)
1988/83	42.9		−0.2
	(4.5) ***	1	(0.1) *
1984	107.5		−0.4
	(15.9) ***	2.5	(0.2) †
1985	182.7		−0.5
	(8.3) ***	4.3	(0.1) ***
1986	225.7		−0.5
	(16.6) ***	5.3	(0.1) ***
1987	240.5		−0.4
	(24.1) ***	5.6	(0.1) *
1988	256.0		−0.4
	(17.7) ***	6.0	(0.1) ***
1989	329.6		−0.7
	(45.7) ***	7.7	(0.2) ***
1990	281.0		−0.6
	(24.1) ***	6.6	(0.2) ***

Note. Parameter estimates are indicated by a "^" with standard errors in brackets.
[a]Model includes a baseline transmission rate (β_0) and a seasonal factor (β_1).
[b]R_0 corresponds to the ratio of $N\hat{\beta}_0$ estimated for the years 1984 to 1990 to its value for the stable period.
*$P < .05$, ***$P < .001$, †$P < .06$, where P is the level of significance of the parameters.

TABLE 3—Estimation of the Parameters for Model 2[a] for *S. enteritidis* in France from 1978 to 1988

Period	$N\hat{\beta}_0$ ($\hat{\sigma}_0$)	R_0[b]	$\hat{\beta}_1$ ($\hat{\sigma}_1$)	$N\hat{\beta}_2$ ($\hat{\sigma}_2$)	R_2[c]
1978/86	32.9		−0.9	6.1	
	(4.0) ***	1	(0.2) ***	(3.0) *	1
1987	76.8		−1.1	10.9	
	(19.5) ***	2.3	(0.3) **	(26.6) NS	1.2
1988	130.6		−0.9	31.0	
	(10.4) ***	4.0	(0.1) ***	(14.2) †	5.1

Note. Parameter estimates are indicated by a "^" with standard errors in brackets.
[a]Model includes a baseline transmission rate (β_0), a seasonal factor (β_1), and a seasonal laying effect (β_2).
[b]R_0 corresponds to the ratio of $N\hat{\beta}_0$ estimated in 1987 or in 1988 to its value for the stable period.
[c]R_2 corresponds to the ratio of $N\hat{\beta}_2$ estimated in 1987 or in 1988 to its value for the stable period.
*$P < .05$, **$P < .01$, ***$P < .001$, †$P = .057$, NS: $P \geq .1$, where P is the level of significance of the parameters.

It is noteworthy that the ratios R_0, which characterize the pattern of increase of β_0, are very similar between French and US data for the first 2 years of the epidemic. However, seasonal patterns are somewhat different between the two regions, with France exhibiting a more stable seasonal pattern.

Linking the Time Pattern of S. enteritidis *Infection to That of a Seasonal Laying Effect of Free-Range Hens*

In France, the involvement of flocks from small holdings in the epidemic can be tested by using model 2. This model was estimated on the stable period (with autoregressive parameters for the residuals at lags 1, 3, 8, 12, and 24) and then separately in 1987 and 1988. All parameters are clearly significant except for $N\beta_2$, which is nonsignificant in 1987 and only marginally significant in 1988 (Table 3). The seasonal multiplicative effect appears reasonably constant.

Comparison of Tables 1 and 3 shows that there is a fairly linear increase of the baseline parameter β_0 for the years 1987 and 1988 at a similar rate for both compartmental models. As expected, the addition of a parameter in model 2 resulted in a smaller value for the baseline rate $N\beta_0$. For the stable period, model 2 gives a significant improvement over model 1 ($\chi_1^2 = 4.04$, $P = .04$), and the seasonal laying effect parameter $N\beta_2$ is significant. In 1987, $N\beta_2$ is not significant, and the standard error of β_1 is increased. As seen in Figure 2, the increase in 1987 is extremely brutal, and neither model 1 nor 2 fits well. On the other hand, $N\beta_2$ is again significant in 1988. Overall, model 2 gives a reasonable fit.

To better visualize the two components of model 2—the sinusoidal part corresponding to model 1 and that relating to the seasonal laying effect ($g[t]$ in the appendix)—these components, as well as their sum for the stable period, are represented separately in Figure 6, with parameters chosen in accordance with Table 3.

Model 2 gave a poorer fit than model 1 for the stable period with the US data; consequently, the seasonal laying hypothesis does not seem appropriate for the United States.

Impact of Preventive Measures

France. Since May 1989, eggs from flocks found to be infected have been diverted to pasteurization plants or the flock has been destroyed. In 1989, for example, eggs were diverted to pasteurization plants for six flocks while three flocks were destroyed. To try to combat the emergence of *S. enteritidis* and reduce the number of infected birds, a working party composed of health officials and egg producers was set up. Specific recommendations concerning hygiene and poultry health checks were made during 1990. A convention was signed between the French state and the egg producers, whereby the state will compensate the producers for the eradiction of infected flocks and will reimburse costs for bacteriological analyses made in this context.

We have based our evaluation of preventive measures solely on the yearly totals for 1989 and 1990 since they can be globally corrected for bias.

Yearly forecasts of infected cases (as well as their standard errors) were computed with model 2 with parameter values replaced by their respective predictions. The parameter predictions were calculated under the assumption that the linear increase of the baseline rate observed for 1987 and 1988 was continuing. Precisely, predictions for parameters $N\beta_0$, NC_1, and $N\beta_2$ in 1989 and 1990 were obtained using equations corresponding to weighted regression of the estimated values of those parameters for three time periods: the stable period, 1987, and 1988. As expected, the parameter $\hat{\beta}_1$ was stable for both years, and the standard errors of $N\hat{\beta}_0$ and $N\hat{\beta}_2$ increased between 1989 and 1990.

The observed and bias-corrected numbers of isolates in the National Reference Center for 1989 and 1990 are given in Table 4. The bias-corrected figures were obtained after first allowing for a 16% increase in the total number of isolates observed (National Reference Center and laboratories) and then taking 78% of these isolates (see "Material and Methods" above).

In 1989, the bias-corrected number of isolates was higher than the predicted level, assuming a linear increase, and was close to the latter's upper 95% confidence interval. This seems to indicate that *S. enteritidis* contamination was accelerating. However, this acceleration did not appear to continue in 1990, as the predicted total was nearly identical to the bias-corrected value for that year. This points to a limited impact of preventive measures, but there is no evidence that the linear rate of increase was diminishing.

United States. In the United States, a voluntary state quality assurance program for *S. enteritidis* in flocks was implemented in February 1990.[5] Forecasts of $N\beta_0$ were calculated for the year 1990 under the assumption of a linear increase in the baseline parameter (Table 5). In 1990 the estimated value of $N\beta_0$ is below the lower 95% confidence interval for the predicted $N\beta_0$, and there is little overlap between their respective 95% confidence intervals. This indicates a potential effect of the preventive measures.

TABLE 4—Yearly Number of Isolates in the National *Salmonella* Reference Center (Observed and Predicted) for *S. enteritidis* in France for 1989 and 1990

Year	Observed (No Bias Correction)	Observed (Bias Correction)[a]	Predicted[b] (95% CI)
1989	3357	3068	2436 (2076, 2796)
1990	2133	3075	3103 (2378, 3828)

[a]After first allowing for a 16% increase in the total number of isolates observed (National Reference Center and laboratories) and then taking 78% of these isolates.
[b]Predicted by continuing the linear increase observed in 1987 and 1988.

TABLE 5—Comparison of Estimated and Predicted Baseline Parameter (β_0) for *S. enteritidis* in the Middle Atlantic Region of the United States for 1990

Year	Estimated (95% CI)	Predicted[a] (95% CI)
1990	128.1 (233.7, 328.3)	415.3 (384.3, 446.3)

[a]Predicted by continuing the linear increase observed between 1984 and 1989.

Discussion

The value of the model used here, which relies on strongly simplifying hypotheses, is that it summarizes the observed phenomenon with some easily interpretable parameters. A time periodical function similar to ours has been used in a measles transmission study.[10] By attempting to represent the underlying phenomenon, deterministic models lead to predictions that are based on explicit hypotheses. An alternative way to make predictions would be to use time series modeling, as was done for *Salmonella bovismorbificans*.[11] However, this purely statistical approach was found to perform less satisfactorily, and it permits neither time effects such as seasonal laying to be tested nor comparison with other data sources to be made.

Since the French data were reported through a system of passive surveillance, they represent only a fraction of all *Salmonella* infections. This surveillance system is understandably sensitive to any change in reporting practices. To assess the effects of variations in reporting, we examined trends in the proportion of isolates sent to the National Reference Center. Hence, even though we could not estimate a model on monthly data for the years 1989 and 1990, we were able to correct yearly totals for bias and compare the forecasted values to the observed ones (corrected for bias) to determine the impact of the preventive measures. Clearly, if monthly data for 1989 and 1990 could have been used, we could have characterized the effect of preventive measures by directly studying their influence on the values of $N\beta_0$ and $N\beta_2$, as was done with the US data. In the United States, such a blurring due to increased reporting is not thought to have occurred because the increase was limited to particular serotypes.[7] The observed stabilization of β_0 in 1990 for the US data could thus be attributed to the preventive measures.

Our results support the hypothesis that the observed increase of *S. enteritidis* started in 1987 in France, even though this phenomenon was present as early as 1984 in other European countries (England, Spain) and in the United States. For comparison between the Middle Atlantic region and France, only the simple seasonal model was used because there are fewer small-holding flocks in the Middle Atlantic region. The striking similarity of the rate of increase of β_0 between France and the US Middle Atlantic region suggests that common processes, probably connected with egg production in both countries, were involved in the *S. enteritidis* epidemic.

In France, the potential part played by free-range hen production from small holdings is shown by comparing the simple seasonal model and the model including the time pattern of seasonal laying. Note that the same time pattern is not observed in commercially distributed free-range eggs, which accounted for only 2.7% of the total 1990 production, since these flocks' rearing conditions ensure an uninterrupted laying throughout the year similar to industrial laying. The *S. enteritidis* increase could be due to the importation of infected birds, which are mainly introduced into nonindustrial flocks. From 1987, diversification of strains of hens corresponding to an increase in imports from other European countries was observed. Nevertheless, it is important to point out that, in common with any seasonal effect, this seasonal laying effect could be interpreted as a summer amplitude correction independent of a direct link with hen laying. It would thus be interesting to see whether this particular time pattern is also present in the structure of indicators of *S. enteritidis* infections in other countries where, as in France, free-range hens in small holdings contribute to some extent to the egg production. □

Acknowledgments

This study was supported in part by the French Ministry of Health.

The authors wish to thank P. A. D. Grimont and P. Bouvet of the Pasteur Institute in Paris for providing the French data, and R. V. Tauxe and N. Bean of the Centers for Disease Control and Prevention in Atlanta for providing the US data. This article has benefited from constructive comments made by M. E. Halloran and anonymous referees, whose help is gratefully acknowledged.

References

1. *Salmonella enteritidis* phage type 4: chicken and eggs. *Lancet.* 1988;ii:720–723. Editorial.
2. St Louis ME, Morse DL, Potter ME, et al. The emergence of grade A eggs as a major source of *Salmonella enteritidis* infections: new implications for the control of salmo-

nellosis. *JAMA.* 1988;14:2103–2107.
3. Mandell JL. *Principles and Practices of Infectious Diseases.* New York, NY: John Wiley and Sons; 1985.
4. Hubert B. Surveillance et prévention des salmonelloses en France. *Med Mal Infect.* 1992;22(spécial issue):325–330.
5. Outbreak of *Salmonella enteritidis* infection associated with consumption of raw shell eggs, 1991. *MMWR.* 1992;41:369–372. Editorial.
6. Chalker RB, Blaser MJ. A review of human salmonellosis: III. magnitude of salmonella infection in the United States. *Rev Infect Dis.* 1988;10:111–124.
7. Hargrett-Bean NT, Pavia AT, Tauxe RV. *Salmonella* isolates from humans in the United States, 1984–1986. *MMWR.* 1988; 37:25–31.
8. Anderson RM, May RM. *Infectious Diseases of Humans.* Oxford, England: Oxford University Press; 1991.
9. *SAS/ETS® User's Guide, Version 6.* 1st edition. Cary, NC: SAS Institute, Inc; 1988.
10. Aron J. Multiple attractors in the response to a vaccination program. *Theor Popul Biol.* 1990;38:58–67.
11. Watier L, Richardson S, Hubert B. A time series construction of an alert threshold with application to *S. bovismorbificans* in France. *Stat Med.* 1991;10:1493–1509.

APPENDIX—The Compartmental Susceptible-Infected Model

$$\boxed{NS(t)} \xrightarrow{f(t)} \boxed{NI(t)}, \qquad \boxed{NI(t)} \xrightarrow{\gamma} \boxed{NS(t)}$$

where $S(t)$ denotes the proportion of individuals not infected but susceptible to contract *S. enteritidis* infection at time t;
$I(t)$ denotes the proportion of individuals infected with *S. enteritidis* at time t;
N is the total population size, assumed to be constant, and $f(t)$ and γ are defined as in the text.

Note that $NI(t)$ represents the number of cases of the disease at time t. This diagram leads to differential equation systems:

(1) $\frac{d}{dt}(NI(t)) = f(t)\,NS(t) - \gamma NI(t)$,

(2) $\frac{d}{dt}(NS(t)) = -\frac{d}{dt}(NI(t))$ (because the total population is assumed to be constant).

Under the hypothesis that the susceptible population is much larger than the infected one, we can suppose that $NS(t) \approx N$, and then $Nf(t)$ in (1) corresponds to the appearance of new cases at time t. In other words, $Nf(t)$ models the time evolution of the incidence of the disease. The system leads (after division of (1) by N) to the differential equation

$$\frac{d}{dt}I(t) + \frac{I}{\gamma(t)} = f(t),$$ which can be integrated to

$$I(t) = e^{-\gamma t}\int_0^t e^{\gamma s} f(s)\,ds + I(0)\,e^{-\gamma t}.$$

If we work with prevalence data, we use the solution $NI(t)$ to model the data. In our case, we dispose of the number of new isolates $O(t)$ at time t. So we directly fit $Nf(t)$ to our data.

Model 1: $f(t) = \beta_0\left(1 + \beta_1 \sin\frac{2\pi t}{12}\right)$, and we fit:

$$O(t) = N\beta_0 + NC_1 \sin\frac{2\pi t}{12} + u_1(t),$$

where $C_1 = \beta_0\beta_1$ and $u_1(t)$ represents an autocorrelated error term with zero mean.

Model 2: $f(t) = \beta_0\left(1 + \beta_1 \sin\frac{2\pi t}{12}\right) + \beta_2\, g(t)$,

with $g(t) = \begin{cases} 1 & \text{between } a + 12i \text{ and } b + 12i \text{ inclusive, } i \geq 0 \\ 0 & \text{otherwise,} \end{cases}$

and we fit:

$$O(t) = N\beta_0 + NC_1 \sin\frac{2\pi t}{12} + N\beta_2\, g(t) + u_2(t),$$

where $C_1 = \beta_0\beta_1$ and $u_2(t)$ represents an autocorrelated error term with zero mean.

Note that the estimated parameters are $N\beta_0$ and $N\beta_2$. Because we made the hypothesis of a constant population N, we can interpret the evolution of the baseline transmission rate β_0 and of the seasonal laying effect β_2.

The Health Effects of Swimming at Sydney Beaches

ABSTRACT

Objectives. The purpose of the study was to determine the health risks of swimming at ocean beaches in Sydney, Australia.

Methods. From people attending 12 Sydney beaches in the period from December 5, 1989 to February 26, 1990, we recruited a cohort of 8413 adults who agreed to participate in this study. Of these, 4424 were excluded either because they had been swimming in the previous 5 days or because they reported a current illness. Of the remainder, 2839 successfully completed a follow-up telephone interview conducted within 10 days after recruitment. We recorded reported respiratory, gastrointestinal, eye, and ear symptoms and fever that occurred within the 10 days between initial interview on the beach and the follow-up interview.

Results. A total of 683 participants (24.0%) reported experiencing symptoms in the 10 days following initial interview. Of these, 435 (63.7%) reported respiratory symptoms. Swimmers were almost twice as likely as nonswimmers to report symptoms. There was a linear relationship between water pollution and all reported symptoms with the exception of gastrointestinal complaints.

Conclusions. Swimmers at Sydney ocean beaches are more likely to report respiratory, ear, and eye symptoms than beachgoers who do not swim. The incidence of these symptoms increases slightly with increasing levels of pollution. (*Am J Public Health.* 1993;83:1701–1706)

Stephen J. Corbett, MRCGP, MPH, FAFPHM, George L. Rubin, FRACP, FAFPHM, Gregory K. Curry, BAppSc, MStats, David G. Kleinbaum, PhD, and the Sydney Beach Users Study Advisory Group

Introduction

The disposal of human sewage through ocean outfalls adjacent to recreational bathing areas and, before that, into Sydney Harbour has been a contentious issue since the early years of the colony in New South Wales.[1] Apart from the aesthetic objections, there is a long-standing concern that bathing in polluted waters poses a risk to the health of the estimated 3 million people who use these beaches annually.[2]

Seventy percent of Sydney's sewage is discharged through three ocean outfalls at North Head, Bondi, and Malabar. Each outfall releases primary treated sewage into coastal waters, 150 to 650 ML per day in dry weather and 700 to 2400 ML per day in wet weather.[3] The sewage system is designed to overflow into the storm-water system at times of high rainfall to cope with this large increase in volume in wet weather. Fecal contamination of swimming beaches can occur when the sewage plume from these outfalls, under the influence of winds or currents, moves onto the beach and via discharge from storm-water drains, many of which discharge directly onto beaches.

Studies at ocean and freshwater sites in the United States,[4–6] Canada,[7,8] Egypt,[9] and Israel[10] have shown a variable relationship between gastrointestinal, respiratory, ear, and skin symptoms and indices of sewage pollution of bathing water. The results of these studies are not necessarily applicable to Sydney's beaches because of differences in dilutive factors, water temperature, ocean currents, and method of sewage treatment and their potential effects on the survival of pathogenic organisms[11] in bathing waters.

In the summer of 1989, the New South Wales Health Department released the results of a survey of bathing water quality that showed that one third of all samples tested at many popular Sydney beaches failed to meet existing water quality criteria.[12] A wave of media and community protest followed. Bodies such as the Australian Surf Lifesaving Association, whose members are the very icons of Australian beach culture, added their voices to the call for an urgent review of bathing water quality in Sydney. Physicians practicing in beachside suburbs reported an increase in the incidence of ear infections, gastroenteritis, and other viral illnesses that could be attributed to ocean bathing.[13] The desertion of beaches that followed these reports had major economic implications for local businesses and for Sydney's reputation as an international tourist destination. Against this background the Sydney Beach Users Study was commissioned, with a brief to investigate the relationship between ocean swimming at Sydney's beaches and illness. An advisory committee comprising representatives from the state government and consumer groups was formed to oversee the planning and execution of the

Stephen J. Corbett, George L. Rubin, and Gregory K. Curry are with the Epidemiology and Health Services Evaluation Branch, New South Wales Health Department, Sydney, Australia. David G. Kleinbaum is with the Division of Epidemiology, Emory University School of Public Health, Atlanta, Ga. For a list of the members of the Sydney Beach Users Advisory Group, see the Acknowledgments.

Requests for reprints should be sent to Stephen J. Corbett, MRCGP, MPH, FAFPHM, Epidemiology Branch, New South Wales Health Department, Locked Mail Bag 961, North Sydney 2059, NSW Australia.

This paper was accepted May 18, 1993.

TABLE 1—Geometric Mean Bacterial Counts of Beach Water Samples, by Organism and Time of Collection, Sydney Beach Users Study, 1989/90

	Colony-Forming Units per 100 mL			No. Samples Passed[a]	No. Samples Failed[a]	Total No. Samples[b]
	Median	Maximum	Mean			
Fecal coliforms						
Morning	26.2	20000.0	591.4	191	42	233
Afternoon	15.9	25570.9	389.1	203	25	228
Fecal streptococci						
Morning	16.4	1356.5	65.9	...[a]	...[a]	229
Afternoon	11.5	1263.3	42.7	...[a]	...[a]	223

[a]For coliforms, a sample failed if the geometric mean count exceeded 300 colony-forming units per 100 mL or if any one reading was greater than 2000 colony-forming units per 100 mL. There is currently no recreational-water quality standard for fecal streptococci in New South Wales.
[b]A total of 482 samples were tested; 21 tests for fecal coliforms and 30 tests for fecal streptococci did not yield usable results.

study. This paper reports the results of this epidemiological study.

Methods

Subjects

Subjects for this cohort study were selected from persons attending 12 designated Sydney beaches on a total of 41 sampling days including weekends, public holidays, and one randomly selected weekday between December 5, 1989, and February 26, 1990. We selected the 12 most popular beaches in the metropolitan area, as assessed by the Surf Lifesaving Association, to recruit subjects.

On each of the study days, beachgoers at 6 of the 12 beaches were approached to obtain consent to participate in the study. The 6 beaches selected on any day included 2 each from the northern, eastern, and southern areas. We used a quota sampling technique for selection of subjects at individual beaches, setting equal quotas for each beach on each sampling day. Each beach was divided into three sectors, defined by the position of the swimming area flags erected by the lifesavers (lifeguards). The allotted quota for the beach was divided equally among the three sectors. Trained interviewers recruited subjects, starting at the center of each sector and moving in a clockwise fashion until the quota for that sector had been reached. To minimize the inclusion in the study of family or social groups, we specified that potential subjects be at least 3 m apart.

To quality for inclusion, a respondent had to be 15 years old or older and had to agree to both an initial beach interview and a follow-up telephone interview. We excluded children younger than 15 years of age because of the problems involved in obtaining consent from minors for biomedical research. At the beach interview the interviewer introduced the respondent to the survey, collected basic demographic data, and asked whether the respondent had swum anywhere in the previous 5 days or had any condition that precluded swimming on the day of the interview.

Subjects were excluded from the study if they reported having been swimming in the 5 days prior to the beach interview or having an illness that prevented them from swimming on the interview day.

Interviewers telephoned study participants 7 to 10 days after the initial interview to elicit details of symptoms of illness in the previous week, attendance at a doctor's office, absence from work, and whether the subject had been swimming at any time between the initial and follow-up interviews. Respondents were asked whether they had experienced vomiting; diarrhea; cough, cold, fever, or symptoms suggestive of flu; or ear or eye infections since the day of the beach interview.

Field interviewing, clerical checking, data entry, and initial computer editing were performed by AGB McNair, a division of AGB Research Australia Pty Ltd under contract to the New South Wales Health Department.

Exposure Assessment

We defined swimming as immersion of the face and head in water. At follow-up the interviewer asked whether the subject had swum on the day of the beach interview and asked the subject to estimate the amount of time spent in the water.

Water samples were taken on the same days and at the same beaches on which interviews were conducted. Health department surveyors collected morning and afternoon samples on each day from sites approximately at the midpoint of each sector of the beach. As nearly as could practicably be achieved, interviewing of respondents at selected beaches was done in the period in which water samples were taken. A standard protocol for the collection of 250 mL of water at a depth of 30 to 45 cm was used. Samples were refrigerated at 4°C and transported for analysis at the Division of Analytical Laboratories of the New South Wales Health Department. Recommended standard methods were used to count fecal coliforms[14,15] and fecal streptococci.[16]

We defined fecal contamination of beachwater as low if there was compliance with the current New South Wales Department of Health standard—geometric mean coliform count not exceeding 300 colony-forming units per 100 mL and no single sample exceeding 2000 colony-forming units per 100 mL—and high if these levels were exceeded.

Statistical Methods

We used the Statistical Analysis System (SAS)[17] for tabulations and subsequent statistical modeling. Unconditional logistic regression, using both the CATMOD procedure in SAS (Version 6.03) and the MULTLR program, was used to model the main effects and covariates on the probability of reporting symptoms.[18] Point and interval estimates in the presence of interaction were performed according to the methods of Kleinbaum et al.[19]

We assessed the health effects of swimming in polluted water by fitting a series of logistic regression models to the data to adjust—for age, sex, and a history of subsequent swimming—estimates of odds of reporting specific symptoms in swimmers compared with nonswimmers. Two variables denoted exposure to polluted water in these models: swimming duration and a measure of bacterial count in the water. The bacterial count was derived from the geometric or arithmetic mean organism count or from the maximum organism count of all water samples collected on that beach on the day of the beach interview.

These logistic regression models were fitted separately for each outcome, that is, (1) any reported symptoms, (2) respiratory symptoms, (3) fever, (4) eye symptoms, (5) ear symptoms, and (6) gastrointestinal symptoms.

Results

Water Sampling

Of 482 individual water samples tested, 461 yielded usable results. Of these, 67 (14.5%) failed to meet the water quality criteria set by the New South Wales Department of Health. Summary results of water testing for both fecal contamination indices measured, coliforms and fecal streptococci, are presented in Table 1. There was a strong correlation between a "pass" on the morning test and a "pass" in the afternoon test. (A water sample was categorized as a "pass" if the geometric mean fecal coliform count was below 300 colony-forming units per 100 mL.) On days when the water failed the morning test, there was a 50% chance of its also failing the afternoon test.

Exposure and Reported Symptoms

Of the 9650 persons who were approached on the selected beaches, 8413 (87.2%) agreed to participate. Of these, we excluded 325 (3.9%) because they had illnesses that prevented them from swimming on that day and 4099 (48.7%) because they reported that they had been swimming in the previous 5 days. Of the remaining 3989 persons, we were able to contact 2968 (74.4%) by telephone for a follow-up interview. Young males were the most likely not to respond at the follow-up interview. The age and sex of those refusing to participate, those excluded because of prior swimming or illness, and those enrolled in the study were similar (Table 2).

We excluded 12 subjects from the analysis because of missing age and sex data and a further 117 subjects because of missing data on pollution levels on the day of beach interview. This left 2839 subjects available for analysis. Of this group, 915 (32.2%) reported that they had not swum on the day of the beach interview. Only 303 (10.7%) of the subjects swam on days on which high pollution levels were recorded.

A total of 683 subjects (24.0%) reported experiencing symptoms in the previous week; of these, 435 (63.7%) reported respiratory symptoms. The prevalence of reported fever and respiratory, eye, ear, and other symptoms increased with increasing bacterial counts measured on the day of initial interview. Gastrointestinal symptoms, which were reported by 4.1% of the subjects, did not increase with increasing counts of fecal bacteria (Table 3).

The prevalence of respiratory symptoms in young people aged 15 through 25 years was high, irrespective of swimming status or pollution level. Only in older people did the incidence of respiratory symptoms rise with increasing levels of pollution. The prevalence of other reported symptoms did not vary significantly across age groups (Table 4).

Swimmers were almost twice as likely as nonswimmers to report symptoms (odds ratio [OR] = 1.9; 95% confidence interval [CI] = 1.4, 2.4), after adjustments were made for age and sex. Symptoms were reported less frequently by people who swam between the original and follow-up interviews (Table 5). There was no evidence that swimmers were deterred from entering the water because it was or appeared to be contaminated.

We assessed the effect of increasing pollution levels by comparing the estimated odds ratios at different levels of pollution for swimmers vs nonswimmers. As the counts of fecal bacteria measured on the day of original interview rose, swimmers were more likely to report all symptoms except gastrointestinal symptoms

TABLE 2—Persons Approached to Participate in the Sydney Beach Users Survey, by Age, Sex, Area of Residence, and Study Status

		Excluded			
	Initial Nonresponse	Swam Previously	Illness	Total Enrolled	Total Contacted at Follow-up
Age					
15–19 y	184	933	61	930	649
20–49 y	871	2734	217	2803	2109
50+ y	130	381	44	240	200
Unknown	72	51	3	16	10
Total	1237	4099	325	3989	2968
Sex					
Male	591	2455	140	2041	1436
Female	580	1608	182	1944	1529
Unknown	66	36	3	4	3
Total	1237	4099	325	3989	2968
Place of residence					
Sydney	...	...	...	3554	2672
Other NSW	...	...	...	254	188
Other	...	...	...	94	52
Unknown	...	...	...	87	56
Total	...	...	...	3989	2968

TABLE 3—Percentages[a] of Beachgoers Reporting Symptoms, by Swimming Status and Pollution Level, Sydney Beach Users Study, 1989/90

	Swimming Status/Pollution Level, %			
	Did Not Swim (n = 915)	Swam, Low Pollution[b] (n = 1770)	Swam, High Pollution[c] (n = 154)	Total Sample (n = 2839)
Vomiting	0.9	1.0	0.6	0.9
Diarrhea	2.2	3.7	3.2	3.2
Cough, cold, flu	10.2	17.3	23.4	15.3
Ear infection	1.3	3.9	5.8	3.2
Eye infection	1.0	2.4	3.9	2.0
Fever	1.1	1.8	5.2	1.7
Other conditions	4.7	8.0	13.0	7.2
Any condition reported	16.5	26.9	35.7	24.0
Attended a doctor (any condition)	3.5	4.3	8.4	4.3
Took time off work (any condition)	2.8	4.6	6.5	4.0

Note. Symptoms were self-reported; respondents could report multiple symptoms.
[a]Percentages are based on column totals.
[b]Geometric mean fecal coliform count of less than 300 colony-forming units per 100 mL recorded on the day of initial interview.
[c]Geometric mean fecal coliform count of more than 300 colony-forming units per 100 mL recorded on the day of initial interview.

TABLE 4—Numbers and Percentages of Persons Who Reported Symptoms, by Age, Swimming Status, and Pollution Level, Sydney Beach Users Study, 1989/90

	Swimming Status/Pollution Level															
	Did Not Swim (n = 915)				Swam, Low Pollution[a] (n = 1770)				Swam, High Pollution[b] (n = 154)				Total Sample (n = 2839)			
	15–24 y	25–39 y	40+ y	All Ages	15–24 y	25–39 y	40+ y	All Ages	15–24 y	25–39 y	40+ y	All Ages	15–24 y	25–39 y	40+ y	All Ages
Gastrointestinal symptoms	11 (3.4)	6 (1.8)	7 (3.1)	24 (2.6)	35 (4.5)	30 (4.7)	13 (3.7)	78 (4.4)	1 (1.4)	3 (5.9)	1 (3.2)	5 (3.2)	47 (4.0)	39 (3.6)	20 (3.3)	107 (3.8)
Cough, cold, flu	56 (17.2)	23 (6.3)	14 (6.2)	93 (10.1)	161 (20.9)	113 (17.6)	32 (9.0)	306 (17.3)	17 (23.3)	12 (23.5)	7 (22.6)	36 (23.2)	234 (20.0)	148 (13.6)	53 (8.7)	435 (15.3)
Ear infection	5 (1.5)	3 (0.8)	4 (1.8)	12 (1.3)	38 (4.9)	18 (2.7)	13 (3.7)	69 (3.9)	5 (6.9)	2 (3.9)	2 (6.5)	9 (5.8)	48 (4.1)	23 (2.1)	19 (3.1)	57 (2.0)
Eye infection	5 (1.5)	4 (1.1)	0 ...	9 (1.0)	15 (1.9)	22 (3.3)	5 (1.4)	42 (2.4)	1 (1.4)	5 (9.6)	0 ...	6 (3.9)	31 (2.6)	31 (2.8)	5 (0.8)	90 (3.2)
Fever	5 (1.5)	1 (0.3)	4 (1.8)	10 (1.1)	15 (1.9)	12 (1.8)	4 (1.1)	31 (1.8)	3 (4.1)	3 (5.8)	2 (6.7)	8 (5.2)	16 (1.4)	16 (1.5)	10 (1.7)	49 (1.7)

Note. Numbers in parentheses are percentages of total n's in the relevant age group and swimming/pollution category (n's for age groups not shown).
[a]Geometric mean fecal coliform count of less than 300 colony-forming units per 100 mL recorded on the day of initial interview.
[b]Geometric mean fecal coliform count of more than 300 colony-forming units per 100 mL recorded on the day of initial interview.

TABLE 5—Odds Ratios (ORs)[a] of Swimmers' Reporting Health Problems, Sydney Beach Users Study, 1989/90

	Swam Once Only		Swam More Than Once		All Swimmers	
	OR	95% CI	OR	95% CI	OR	95% CI
Any symptom	2.6	1.8, 3.8	1.6	1.2, 2.2	1.9	1.4, 2.4
Cough	2.4	1.6, 3.8	1.6	1.1, 2.3	1.9	1.4, 2.6
Ear symptoms	6.1	1.7, 21.8	2.8	1.3, 6.2	3.6	1.8, 7.2
Eye symptoms	4.4	0.9, 21.3	1.1	0.4, 2.9	1.7	0.7, 4.3
Fever	1.8	0.5, 6.0	1.1	0.4, 3.2	1.4	0.5, 3.7
Any gastrointestinal symptom	1.8	0.7, 4.6	1.3	0.7, 2.6	1.5	0.8, 2.7

Note. CI = confidence interval.
[a]Adjusted for age, sex, swimming duration, and geometric mean fecal coliform count; the reference category is nonswimmers.

TABLE 6—Odds Ratios (ORs)[a] of Swimmers' Reporting Health Problems, by Specific Symptom and Pollution Level, Sydney Beach Users Study, 1989/90

	Pollution Level							
	10–300 cfu/100 mL		300–1000 cfu/100 mL		1000–3000 cfu/100 mL		>3000 cfu/100 mL	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Any symptom	2.9	1.7, 5.1	3.8	2.1, 7.1	5.2	1.7, 16.0	5.9	3.0, 11.5
Cough	2.4	1.5, 3.8	2.0	0.9, 4.4	4.2	1.2, 14.6	6.9	3.3, 14.1
Ear symptoms	4.3	1.1, 16.2	6.6	1.7, 43.2	8.5	0.8, 97.6	7.4	1.3, 43.3
Eye symptoms	6.3	1.3, 30.8	9.7	1.5, 63.7	8.7	1.0, 72.8		...[b]
Fever	2.1	0.6, 7.0	4.7	1.0, 22.5	9.0	1.9, 43.5		...[b]
Any gastrointestinal symptom	4.6	1.9, 4.9	3.1	0.7, 13.0	3.4	0.7, 18.0		...[b]

Note. Results are for swimmers who swam only on the day of the original interview and who swam for longer than 30 minutes. CI = confidence interval; cfu = colony-forming units of fecal coliforms.
[a]Adjusted for age, sex, and swimming duration; the reference category is nonswimmers.
[b]Highest category > 1000 cfu/100 mL.

(Table 6). We plotted the estimated odds ratios of swimmers' reporting symptoms (within the group of swimmers only, using the lowest pollution level as the reference group) against pollution level (Figure 1). The linear relationship between pollution level and the odds of reporting symptoms can be expressed mathematically by the formula OR = 1 + 0.0000921 ln (Geometric Mean Coliform Count).

Substituting the arithmetic mean for the geometric mean coliform count in the logistic model slightly improved the fit of this model to the data ($\chi^2 = 2.18$, $P < .25$). Counts of fecal streptococci were worse predictors of the odds of reporting symptoms than were counts of fecal coliforms.

We assessed the effect of swimming duration on health by comparing information on those who swam for less than 30 minutes and those who swam for more than 30 minutes, but only in those subjects who did not swim between interviews. Those who swam for longer than 30 minutes were 4.6 times more likely to develop gastrointestinal symptoms than were nonswimmers or those who swam for less than 30 minutes (Table 7).

Discussion

The cohort design used in this study is an adaptation of that used in similar studies in other countries.[6–9] The design has the explicit aim of isolating exposure to a particular day so that the measurements of beach pollution on that day can

be used as a proxy measure of the dose of potentially pathogenic microorganisms. We attempted to improve on this design by increasing the frequency of pollution measurement on study days and by including data on swimming duration in the follow-up questionnaire. The exclusion of children, done primarily for ethical reasons, reduces reporter bias but limits the applicability of our results to people younger than 15 years.

There are at least two explanations for the lower prevalence of reported symptoms in those who swam between interviews. In some people, illness may have prevented swimming between the day of the original interview and the follow-up call. Conversely, those people who did swim subsequently may have included regular swimmers who have some immunity to pathogenic microorganisms in ocean water.

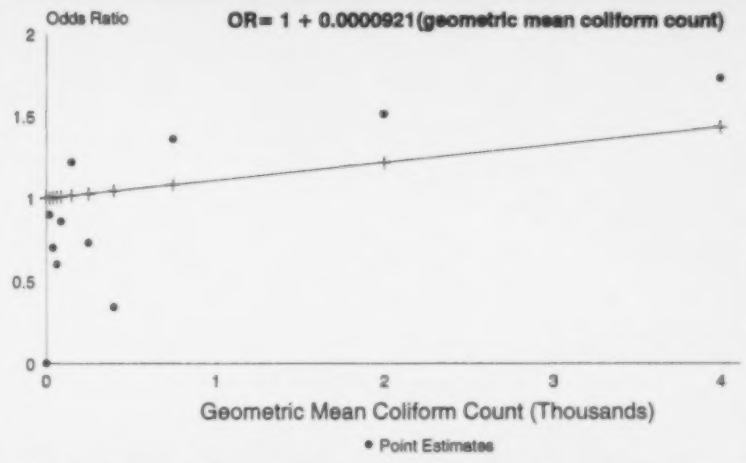
Our results provide evidence that the risk of swimming-related illness increases with increasing pollution levels. The estimates of the magnitude of the slope of the relationship between pollution levels and all symptoms are similar to those obtained elsewhere.[8] The assumption of a linear relationship between illness and pollution levels may have overestimated the risk of illness at lower pollution levels. If a minimum infective dose of microorganisms acquired while swimming is required to cause illness then we should see a threshold pollution level below which illness did not occur. Our data suggest that such a threshold may in fact be present (see Figure 1).

The increases in risk of respiratory, ear, and eye symptoms account wholly for the increases in illness observed. Enteroviruses, the commonest viruses present in sewage effluent, can cause respiratory symptoms.[20] They also persist in marine sediments and water for many months.[21] In a pilot study of viruses and pathogenic bacteria in waters at two Sydney beaches,[22] 2 of 10 samples taken from bathing areas contained enteroviruses; both vaccine-strain polioviruses. However, enteroviruses were isolated in 63% of samples collected from sewage effluent or the sewage plume. All counts of fecal indicator organisms in bathing areas in this study were below the current New South Wales standard, and it may therefore have underestimated the numbers of viruses present in polluted conditions.

Over the 41 days of data collection the beaches were cleaner than usual, thereby limiting the precision of our estimates of the odds of pollution-related symptoms at high levels of pollution. Of the water samples tested in our study, 14% exceeded the current New South Wales standard. Daily surveillance of ocean water quality at seven Sydney beaches between 1983 and 1987[3] showed that five of these beaches exceeded the current standard on more than 30% of occasions in dry weather. In wet weather, six beaches failed more than 50% of the tests.

Our data do not support the contention that minor barotrauma associated with surf swimming is the cause of an increase in reported respiratory symptoms in bathers. If this were true, we would expect an increase in reported respiratory symptoms with increasing swimming duration. We found no such increase.

Reported gastrointestinal symptoms showed no increase with increasing pollution levels. The prevalence of gastrointestinal symptoms (4.1%) was similar to that reported in other studies.[6-9] A dose–response relationship between indices of bacterial pollution and gastrointestinal symptoms has been an inconsistent finding and has been most clearly demonstrated in children, who were excluded from our study. Some of these studies did not fully adjust for confounders such as age and reporter and interviewer bias, and they may have overestimated the magnitude of the dose–response effect.[23] In the study design we used self-reporting of symptoms and exclusion of children to minimize reporter bias. We did not use



Note. ORs are adjusted for age, sex, and swimming duration; swimmers in water with 10 colony-forming units of fecal coliforms per 100 mL were the reference category.

FIGURE 1—Odds ratio (OR) of swimmers' reporting health problems, by geometric mean coliform count, Sydney Beach Users Study, 1989/90.

TABLE 7—Odds Ratios (ORs)[a] of Swimmers' Reporting Health Problems, by Specific Symptom and Swimming Duration, Sydney Beach Users Study, 1989/90

	Swimming Duration			
	< 30 min		> 30 min	
	OR	95% CI	OR	95% CI
Any symptom	1.9	1.3, 2.8	2.9	1.7, 5.1
Cough	1.8	1.1, 2.8	2.4	1.5, 3.8
Ear symptoms	3.7	1.0, 13.5	4.3	1.1, 16.4
Eye symptoms	6.9	1.5, 31.4	6.3	1.3, 30.8
Fever	2.5	0.8, 7.4	2.1	0.6, 7.0
Any gastrointestinal symptom	1.0	0.4, 3.0	4.6	1.9, 10.9

Note. Results are for swimmers who swam only on the day of the original interview and who were exposed to pollution levels of 10–300 colony-forming units of fecal coliforms per mL.
[a]Adjusted for age and sex; the reference category is nonswimmers.

any clinical confirmation of reported symptoms. Measurement error of this kind would be nondifferential and would therefore bias our estimate toward not showing an effect. There was some evidence of an increase in risk of reporting gastrointestinal illness with increased swimming duration. Swimming duration may be a proxy measure of the volume of ingested seawater, which may directly cause symptoms.

The results of this study suggest that, at Sydney's beaches, fecal coliforms are marginally better predictors of reported symptoms than fecal streptococci. These results are at odds with other studies in which fecal coliforms have been poorer predictors of adverse health outcomes than either streptococci or staphylococci. This discrepancy may be explained by the different relative survival of these microorganisms in effluent that is not chlorinated and in oceans that are warmer than those around North America.

A technical issue prominent in the public debate on ocean water quality was the choice of an appropriate summary measure of water quality. Bacterial counts in water samples are inherently variable and concurrent samples from the same site can differ by orders of magnitude. The geometric mean of serial measurements is traditionally used as a summary index of pollution. Criticism of the use of a geometric mean of bacterial counts has focused on the apparent "leveling" of the data,[1] thereby understanding the true level of pollution. Furthermore, use of the geometric mean count rather than, say, the arithmetic mean or the daily maximum may introduce bias.[24]

To address these concerns we substituted the arithmetic mean and the daily maximum for the geometric mean coliform in the logistic models. There was a small improvement in the fit of the data when the arithmetic mean was used, and this summary measure is therefore preferable. It also has the advantage of being more easily comprehended by the public.

In conclusion, we found a consistent increase in reported illness with increasing pollution levels of all symptom categories except gastrointestinal symptoms. This dose–response relationship provides a quantitative basis for developing a health-based standard for ocean water quality on Sydney's beaches. □

Acknowledgments

The Sydney Beach Users Study Advisory Group consisted of (in alphabetical order) Mr Paul Albertini, New South Wales Surf Lifesaving Association; Dr Syd Bell, Eastern Sydney Area Health Service; Mr Andrew Bernard, Division of Analytical Laboratories, New South Wales Department of Health; Dr Charles Bridges-Webb, Royal Australian College of General Practitioners; Dr David Campbell, Northern Sydney Area Health Service; Mr Ernie Davis, New South Wales Surf Lifesaving Association; Ms Nancy Ealer, Senior Policy Adviser to the Minister for the Environment; Dr Julian Gold, Eastern Sydney Area Health Service; Mr Richard Gosden, Greenpeace Australia Ltd; Dr Roderick Kennedy, Central Coast Area Health Service; Dr Peter Macdonald and Dr Colin MacLeod, Central Sydney Area Health Service; Dr Paul Morad, New South Wales Surf Lifesaving Association; Dr Aileen Plant, University of Sydney; Dr Robert Reznik, Central Sydney Area Health Service; Dr Louise Rushworth, Epidemiology and Health Services Evaluation Branch, New South Wales Department of Health; Dr Greg Stewart, Southern Sydney Area Health Service; and Dr David Wilcox, Metropolitan Water, Sewage, and Drainage Board.

References

1. Bede S. *Toxic Fish and Sewer Surfing.* Sydney, NSW, Australia: Allen & Unwin; 1989.
2. *Report to the Sydney Beach Users Study Group.* New South Wales Surf Lifesaving Association, Sydney, NSW, Australia: 1989.
3. Camp, Dresser and McKee International. Review of Sydney's Beach Protection Programme: Report to the New South Wales Minister for the Environment. Sydney, NSW, Australia: NSW Minister for the Environment; 1989.
4. Cabelli VJ, Dufour AP, McCabe LJ, et al. Swimming-associated gastroenteritis and water quality. *Am J Epidemiol.* 1982;115: 606.
5. Cabelli VJ, Dufour AP, Levin MA, et al. Relationship of microbial indicators to health effects at marine bathing beaches. *Am J Public Health.* 1979;69:690–694.
6. *Ocean Health Study: A Study of the Relationship between Illness and Ocean Beach Water Quality in New Jersey.* Trenton, NJ: New Jersey Department of Health, Division of Occupational and Environmental Health; September 1990.
7. Seyfried PL, Tobin RS, Brown NE, et al. A prospective study of swimming-related illness, I. swimming-associated health risk. *Am J Public Health.* 1985;75:1068–1070.
8. Seyfried PL, Tobin RS, Brown NE, et al. A prospective study of swimming-related illness, II. morbidity and microbiological quality of water. *Am J Public Health.* 1985; 75:1071–1074.
9. El Sharkawi F. The pollution of the beaches of Alexandria due to the discharge of sewage: a case study. *Water Sci Technol.* 1986;18:273–278.
10. Fattal B, Peleg-Olevsky E, Agursky T, et al. The association between seawater pollution as measured by bacterial indicators and morbidity among bathers at Mediterranean bathing beaches of Israel. *Chemosphere.* 1987;16(2/3):565–570.
11. Dufour AP. Bacterial indicators of recreational water quality. *Can J Public Health.* 1984;75:49–56.
12. Bernard AG. The bacteriological quality of tidal bathing waters in Sydney. *Water Sci Technol.* 1989;21:65–69.
13. Mealy E, Craig O. Beachside doctors tell of surf victims. Sydney *Sun-Herald.* March 1989:15.
14. Rose RE, Geldreich EE, Litsky W. *Proceedings of a Symposium on the Recovery of Indicator Organisms Employing Membrane Filters.* Washington, DC: Environmental Protection Agency; 1977:101–105. EPA 600/9-77-024.
15. *American Waterworks Association and Water Pollution Control Council Standard Methods for the Examination of Water and Waste Water.* 17th ed. Washington, DC: American Public Health Association; 1989.
16. *Methods for the Examination of Water and Air: Microbiological Examination of Water by Membrane Filtration.* Sydney, Australia: Standards Association of Australia; 1981. Australian Standard 1095.4.1.8.
17. *SAS/Stat Guide for Personal Computers, Version 6.* Cary, NC: SAS Institute Inc; 1987.
18. Campos FN, Franco EL. A micro-computer program for multiple logistic regression by unconditional and conditional maximum likelihood methods. *Am J Epidemiol.* 1989;129:439–444.
19. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiological Research: Principles and Quantitative Methods.* New York, NY: Van Nostrand Reinhold; 1982:477–491.
20. Melnick JL, Wenner HA, Phillips JL. Enteroviruses. In: Lennette EH, Schmidt NJ, eds. *Diagnostic Procedures for Viral, Rickettsial, and Chlamydial Infections.* Washington, DC: American Public Health Association; 1979:471–534.
21. Goyal SM, Adams WA, O'Malley MA, Lear DW. Human pathogenic viruses at sewage sludge disposal sites in the Middle Atlantic Region. *Appl Environ Microbiol.* 1984;48:758–763.
22. Kueh CSW, Grohmann GS. Recovery of viruses and bacteria in waters off Bondi Beach: a pilot study. *Med J Aust.* 1989;151: 632–638.
23. Lightfoot NE. *A Prospective Study of Swimming Related Illness at Six Freshwater Beaches in Southern Ontario.* Thesis.
24. Seixas NH, Robins TG, Moulton LH. The use of geometric and arithmetic mean exposures in occupational epidemiology. *Am J Ind Med.* 1988;14:465–477.

Cat Scratch Disease in the United States: An Analysis of Three National Databases

Lisa A. Jackson, MD, Bradley A. Perkins, MD, and Jay D. Wenger, MD

ABSTRACT

Objectives. Current knowledge of the epidemiology of cat scratch disease is based primarily on information from case series. We used three national databases to obtain more representative data to determine the incidence and demographics of cat scratch disease.

Methods. Records coded with the diagnosis of cat scratch disease from two hospital discharge databases and an ambulatory care database were analyzed. Costs of diagnostic tests and hospitalization were obtained from a sample of providers and published data.

Results. The incidence of patients discharged from hospitals with a diagnosis of cat scratch disease was between 0.77 and 0.86 per 100 000 population per year. Fifty-five percent of the case patients were 18 years of age or younger. Males accounted for 60% of cases. Incidence varied by season; approximately 60% of case patients were discharged in the months September through January. The estimated incidence of disease in ambulatory patients was 9.3 per 100 000 population per year. On the basis of these rates, we estimated the annual health care cost of the disease to be more than $12 million.

Conclusions. The rates and seasonality of cat scratch disease found in this study were consistent with previous reports. Adults represented a higher percentage of the total than reported in previous case series, suggesting that the disease may affect more adults than previously recognized. (*Am J Public Health.* 1993;83: 1707–1711)

Introduction

Cat scratch disease is classically characterized by persistent regional lymphadenopathy in an area draining lymph from the site of a recent cat scratch. Recent evidence suggests that *Rochalimaea henselae,* a bacterium that has been isolated from patients with bacillary angiomatosis,[1] is associated with cat scratch disease.[2–4] Case series have been the primary source of information on the age distribution, seasonality, and clinical characteristics of cat scratch disease since it was first recognized in the United States in the 1940s.[5] Although these reports have contributed greatly to our understanding of the disease, they are subject to geographic and referral biases, and because they are not population-based, they cannot be used to determine rates of disease. Advances in identifying the etiology of the disease, with the associated implications for improvement in diagnostic capabilities and treatment, emphasize the need to determine the magnitude of the disease and the population primarily affected by it. We used information from two large hospital discharge databases and an ambulatory care database to describe the epidemiologic features and estimate the health care costs of cat scratch disease in the United States.

Methods

Abstracted hospital discharge records from the Commission on Professional and Hospital Activities (CPHA) and the National Hospital Discharge Survey (NHDS) were analyzed. Participation in CPHA is voluntary and the percentage of short-stay hospitals participating varies by state and year. The ratio of the number of discharges from CPHA hospitals in a state in a given year to the total number of hospital discharges for the state (as reported by the American Hospital Association) in that year was used to calculate a weight variable to allow extrapolation to national estimates. Records from 1983 through 1989 with cat scratch disease as any one of as many as 22 discharge diagnoses coded were analyzed.

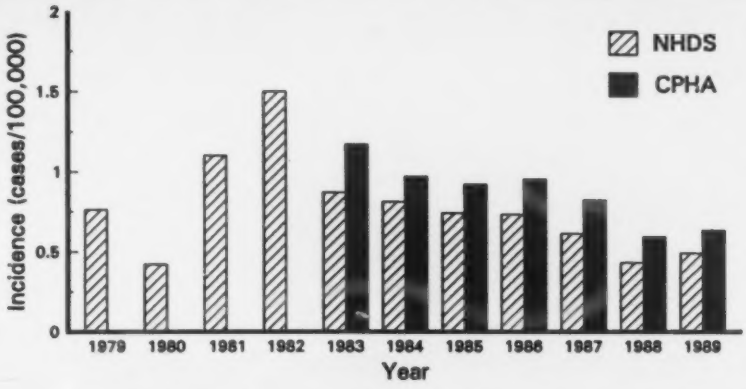
NHDS is a continuous national survey of inpatient utilization of short-stay hospitals conducted by the National Center for Health Statistics (NCHS). Survey data are abstracted from a sample of medical records of patients discharged from nonfederal short-stay hospitals chosen systematically on the basis of hospital size. A weight variable is provided by NCHS to allow extrapolation to national or regional estimates. Records from 1978 through 1989 with cat scratch disease as any one of as many as seven discharge diagnoses coded were analyzed. The relative standard error for the number of records with a diagnosis of cat scratch disease was calculated using information provided by NCHS[6] and was used to determine the 95% confidence intervals (CIs) for the weighted estimates.

The National Ambulatory Medical Care Survey is conducted intermittently by NCHS and provides data from outpatient records selected from a national sample of office-based physicians. A weight variable is provided by NCHS to allow extrapolation to national or regional estimates. Surveys were conducted in 1980,

The authors are with the Division of Bacterial and Mycotic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Ga.

Requests for reprints should be sent to Jay D. Wenger, MD, Mailstop C-09, Centers for Disease Control and Prevention, Atlanta, GA 30333.

This paper was accepted April 9, 1993.



Note. CPHA records from 1983 through 1989 and NHDS records from 1979 through 1989 were analyzed.

FIGURE 1—Incidence of hospitalizations for cat scratch disease, estimated from National Hospital Discharge Survey (NHDS) and Commission on Professional and Hospital Activities (CPHA) data sets, by year.

1981, 1985, and 1989; records with cat scratch disease as any one of the three physician diagnoses coded from these surveys were analyzed. The relative standard error could be calculated from the information provided by NCHS records only for 1985; this figure was used to determine the 95% confidence interval of weighted estimates for that year.

All diagnoses and procedures were coded with *International Classification of Diseases, 9th Revision: Clinical Modification* (ICD-9-CM) codes.[7] Code 078.3 (cat-scratch disease) was used to extract records with the diagnosis of cat scratch disease, and codes 40.0 (incision of lymphatic structures), 40.1 (diagnostic procedures on lymphatic structures), 40.11 (biopsy of lymphatic structure), 40.19 (other diagnostic procedures on lymphatic structures), 40.2 (simple excision of lymphatic structure), 40.21 (excision of deep cervical lymph node), 40.23 (excision of axillary lymph node), 40.24 (excision of inguinal lymph node), 40.29 (simple excision of other lymphatic structure), and 40.3 (regional lymph node excision) were used to extract records with lymph node procedures (excluding excision of internal mammary lymph nodes and radical excision or dissection of any lymph node).

Population estimates were obtained from the US Bureau of the Census.[8] All rates are expressed as cases per 100 000 population per year. Regions listed correspond to those described by the Bureau of the Census. All calculations were performed after records were weighted as described above.

Cost estimates for diagnostic tests and invasive procedures were based on the average price quoted by a sample of providers in the Atlanta metropolitan area in 1992. The cost of an office visit,[9] inpatient day at a community hospital,[10] and hospital consultation[9] were obtained from published data and were adjusted to 1992 dollars at a rate of increase of 6% per year.[11]

Results

Hospitalized Patients

Cat scratch disease was listed as a diagnosis in 2828 records reported in CPHA from 1983 through 1989 and 95 records reported in NHDS from 1979 through 1989; these records represented, in each database, less than 0.01% of the total number of records. After weighting, we estimated averages of 2060 (CPHA) and 1787 (NHDS) (95% CI = 922, 2652) discharges per year of patients with a diagnosis of cat scratch disease. The estimates of the annual incidence of patients discharged with a diagnosis of cat scratch disease were 0.86 (CPHA) and 0.77 (NHDS) per 100 000 population. The incidence estimates varied from 0.59 per 100 000 population in 1988 to 1.2 per 100 000 population in 1983 for CPHA and from 0.42 per 100 000 population in 1980 to 1.5 per 100 000 population in 1982 from NHDS (Figure 1).

The age and sex distributions, regional rates, and seasonal patterns of cases derived from the CPHA and NHDS data sets were similar. Both sets showed a higher proportion of cases among children (persons 18 years of age and younger) than adults. Forty-five percent of CPHA and 50% of NHDS case patients were younger than 15 years, and 54% of CPHA and 57% of NHDS case patients were 18 years old or younger. The highest incidence in the CPHA data set, 2.0 per 100 000 population, was in the 5 through 14-year-old age group; in the NHDS data set the peak incidence, 2.1 per 100 000, was in children younger than 5 years (Figure 2). Males were disproportionately represented. There was a higher incidence in males than in females in both data sets—1.0 and 0.93 per 100 000 males vs 0.69 and 0.60 per 100 000 females for CPHA and NHDS, respectively. Fifty-eight percent of all case patients in the CPHA data set and 60% of all case patients in the NHDS data set were males. The incidence of cat scratch disease was higher among Whites than Blacks. In the CPHA data set, the average incidence was 0.87 per 100 000 for Whites and 0.53 per 100 000 for Blacks; in the NHDS data set the average incidence was 0.78 per 100 000 for Whites and 0.5 per 100 000 for Blacks. Incidence varied by geographic region; in both data sets, the average annual incidence was lower in the West and higher in the South than in the nation as a whole (Figure 3).

Seasonal variation was observed; 59% of CPHA and 61% of NHDS case patients were admitted in the 5 months from September through January (Figure 4). Seasonality was more pronounced for the Midwest region, in which 64% of CPHA and 76% of NHDS case patients were admitted during those 5 months.

Previously described manifestations of cat scratch disease, including encephalopathy, seizures, thrombocytopenia, erythema nodosum, neuritis (brachial, peripheral, or lumbosacral), hepatitis, and osteomyelitis, were reported in 4.9% of CPHA and 8.5% of NHDS records. No patient had a diagnosis of Parinaud's oculoglandular syndrome. No fatalities were reported in NHDS; one death was recorded in CPHA, for a case-fatality rate of 0.03%.

A lymph node procedure—incision, excision, or biopsy—was recorded for 56% of CPHA and 56% of NHDS cat scratch disease case patients. Patients with cat scratch disease accounted for 0.8% of all lymph node procedures recorded for all patients in the NHDS data set from 1979 through 1989. This percentage was higher for younger age groups; cat scratch disease patients accounted for 2.7% of all patients younger than 15 years with lymph node procedures and for 7.4% of all 5- through 14-

year-old patients with lymph node procedures. Cat scratch disease patients accounted for 1.3% of all lymph node procedures performed from September through January and for 15.6% of those performed during this period in the 5- through 14-year-old age group. Data on all patients with lymph node procedures in the CPHA data set were not available.

The average duration of hospitalization for case patients in each database was 4 days (ranges: 0–373 days [CPHA] and 0–16 days [NHDS]). Approximately half of the case patients in both sets stayed 2 days or less.

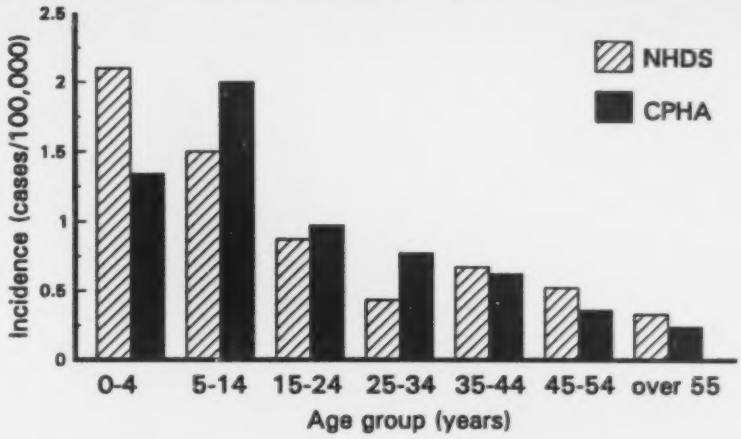
Ambulatory Patients

A total of 10 records with a diagnosis of cat scratch disease were recorded in the National Ambulatory Medical Care Surveys: 1 in 1980, 1 in 1981, 8 in 1985, and 0 in 1989. After weighting, the estimated total number of outpatient cases of cat scratch disease for those years was 87 484, for an average of 22 000 cases per year. Relative standard error could be calculated for 1985 records only; the 95% confidence interval for the estimate of 63 663 cases in 1985 was 0, 147 264. Annual incidence varied from 0 in 1989 to 26.9 per 100 000 population in 1985; the average annual incidence was 9.3 per 100 000 population. All case patients were White; 62% were male. The average age was 22 years (range: 5–33 years), and 54% were younger than 25 years. A fall/winter peak was not seen; only 33% of cases occurred between September and January.

Estimated Cost of Disease

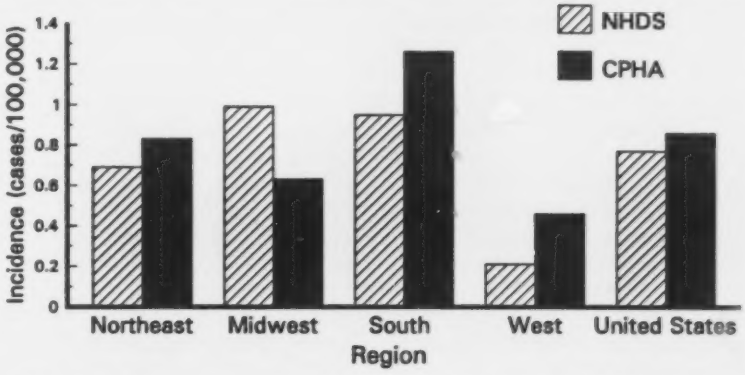
A clinical diagnosis of cat scratch disease requires the fulfillment of three of four criteria: (1) history of cat contact, with the presence of a scratch or primary lesion; (2) positive cat-scratch skin test antigen response; (3) negative laboratory results for other causes of lymphadenopathy; and (4) characteristic pathologic findings of a lymph node.[12,13] The skin test antigen is not standardized, licensed, or widely available; thus, to fulfill the diagnostic criteria a lymph node biopsy and laboratory tests to exclude other causes are usually needed. Therefore, most patients with suspected cat scratch disease would be expected to have a complete blood count, rapid plasma reagin, Monospot, purified protein derivative, and chest radiograph as part of a routine diagnostic evaluation.[14]

We estimated the cost of these tests to be approximately $190. This cost, combined with the cost of an office visit ($45), results in a total cost of approximately $5.2 million for the estimated 22 000 ambulatory cat scratch disease patients per year. Assuming that the cost of one inpatient day in a community hospital ($760) with one physician visit ($40) is incurred for each day of hospitalization and that a surgeon's fee ($400) is billed to the 56% of the hospitalized patients who have lymph node procedures, the estimated cost for the 2000 patients hospitalized annually with cat scratch disease is $6.9 million. Thus the estimated total health care cost of cat scratch disease, in 1992 dollars, is over $12 million per year.



Note. CPHA records from 1983 through 1989 and NHDS records from 1979 through 1989 were analyzed.

FIGURE 2—Mean annual incidence of hospitalizations for cat scratch disease, estimated from National Hospital Discharge Survey (NHDS) and Commission on Professional and Hospital Activities (CPHA) data sets, by age.

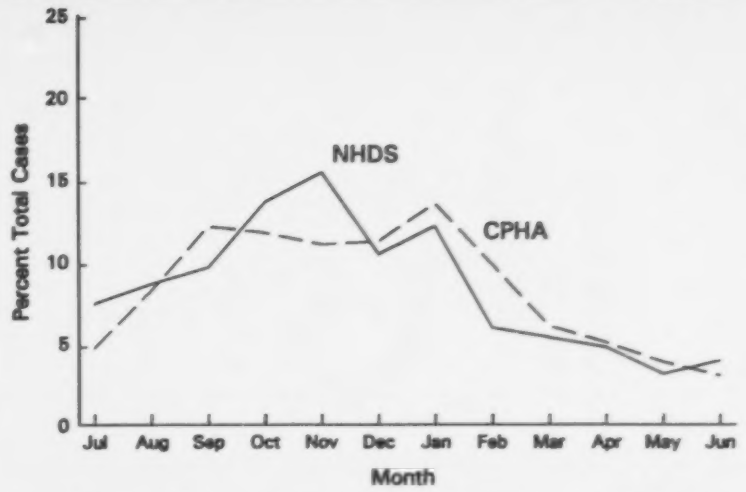


Note. CPHA records from 1983 through 1989 and NHDS records from 1979 through 1989 were analyzed.

FIGURE 3—Mean annual incidence of hospitalizations for cat scratch disease, estimated from National Hospital Discharge Survey (NHDS) and Commission on Professional and Hospital Activities (CPHA) data sets, by region.

Discussion

The rates of cat scratch disease reported here are consistent with previous estimates. The incidence of patients discharged from the hospital with a diagnosis of cat scratch disease, approximately 0.8 per 100 000 population, is similar to the incidence of 0.66 per 100 000 population reported by Carithers from his mail survey of hospitals discharging over 750 pediatric patients annually.[15] The estimate of overall incidence of the disease in ambulatory patients, 9.3 per 100 000 population, is just over 10 times the incidence in hospitalized



Note. CPHA records from 1983 through 1989 and NHDS records from 1979 through 1989 were analyzed.

FIGURE 4—Seasonal variation of hospitalizations for cat scratch disease, estimated from National Hospital Discharge Survey (NHDS) and Commission on Professional and Hospital Activities (CPHA) data sets.

patients. This estimate is similar in magnitude to the rate of 3.3 per 100 000 reported by Margileth[16] on the basis of outpatient studies in the metropolitan Washington, DC, area. Our estimate of the incidence of ambulatory cat scratch disease must be qualified, however, because it was extrapolated from very few cases; additional data, such as might be obtained from population-based surveillance, are needed for a more precise estimate.

The fall/winter peak in incidence noted in the CPHA and NHDS data sets is also congruous with previous reports.[15,17–19] The lack of a finding of seasonality in ambulatory cases is probably due to the small number of records analyzed.

Most investigators, such as Wlodaver and Vorse[20] and Carithers,[15] have found equal rates of disease for both sexes. Margileth,[17] however, in a review of 1252 patients, noted a significantly higher percentage of cases (58%) among males. The predominance of males found in this analysis was noted consistently in all three databases and is therefore likely to reflect a true difference rather than an artifact of data collection. Previous case series have not described the racial distribution of cat scratch disease case patients; however, we found a higher incidence in Whites than in Blacks. Additional investigation of risk factors for cat scratch disease may be useful to explore possible explanations for the gender and racial differences.

The incidence of hospitalizations for cat scratch disease was higher among children than adults, and approximately 54% of CPHA cases and 57% of NHDS cases were in persons 18 years of age or younger. However, children represented a markedly smaller percentage of total cases in both CPHA and NHDS data sets than that noted in most previous reports.[15,17,20,21] Carithers reported that 87% of his 1200 cat scratch disease patients were 18 or younger.[15] Similarly, Margileth, in a study of 1400 patients, noted that 83% were younger than 21,[17] and Wlodaver and Vorse, in a study of 23 patients, noted that 87% were 16 or younger.[20] These case series were reported by pediatricians and the patients were drawn primarily from ambulatory pediatric populations. Therefore, one explanation for the older age distribution found in our analysis is that previous studies were biased toward a younger age group because of the pediatric focus of the investigators. Disease in adults may have been underreported in those case series. This explanation is supported by the findings of Daniels and MacMurray, neither of whom were pediatricians, who described 160 cat scratch disease patients from 27 states and 8 foreign countries and found that only 66% were younger than 30 years of age.[22]

Another explanation for the older age distribution found in CPHA and NHDS records is that a disproportionate number of older patients with cat scratch disease are hospitalized, possibly because the disease is more severe in this age group. This is consistent with the finding of Margileth et al. that only 22% of cat scratch disease patients with severe, systemic disease were younger than 21,[12] compared with 87% of patients with uncomplicated cat scratch disease.[17]

Cat scratch disease is said to be one of the most common causes of chronic regional lymphadenopathy in children.[18,23] In this study, we were unable to compare the incidences of cat scratch disease and other causes of lymphadenopathy in children. However, although children with cat scratch disease accounted for only 3% of lymph node procedures performed on patients younger than 15 years, 16% of lymph node procedures in children aged 5 through 14 years discharged from September through January were in patients with cat scratch disease. It can be concluded from these data that cat scratch disease does cause a significant proportion of lymphadenopathy in children and that a substantial proportion of lymph node procedures performed on children during the fall and winter months are performed because of the disease.

We estimated the annual cost of outpatient evaluation of cat scratch disease to be $5.2 million and the cost of hospitalization to be $6.9 million, for an estimated combined annual cost of the disease of over $12 million. This is meant to be a minimum estimate. The actual cost is likely to be higher, since we did not attempt to include all of the health care costs that may be incurred by patients with cat scratch disease.

Although antibiotics have been suggested by several anecdotal reports to shorten the duration of illness,[23–27] no treatment requiring hospitalization has been proven beneficial to patients with cat scratch disease. Therefore, we assume that the majority of hospitalizations for cat scratch disease are for diagnostic purposes. If this is true, the availability of a reliable, noninvasive diagnostic test would reduce the need for hospitalization. Such a test would also decrease the number of additional tests required for the diagnostic evaluation of ambulatory patients. A recent report indicates that an indirect fluorescent antibody test for detection of humoral antibodies to *Rochalimaea henselae* may be useful for diagnosis of cat scratch disease.[2] Assuming that such a test would allow 90% of hospitalized patients to be treated as outpatients, that this test would be widely available, and that the cost of the test and the cost of

one office visit would be the only costs incurred for outpatients, a test costing $20 per patient would provide a yearly savings of at least $10 million.

This analysis was limited in many reports. Cat scratch disease represented a very small proportion of the total number of records included in the databases analyzed; therefore, the standard error was large and estimates made from these data are not expected to be precise. This problem is unavoidable when researchers seek to determine the incidence of a relatively uncommon disease from surveys that extrapolate from small samples. The diagnosis was obtained from coded records and was not standardized, so some of the patients included in the analysis may not have met standard diagnostic criteria for cat scratch disease. However, the seasonal variation, which was consistent with previous reports, provides some indication that cat scratch disease was correctly coded as a diagnosis for most records. Prospective, population-based surveillance may provide a more accurate estimate of incidence; however, we believe that the results of this analysis accurately reflect the general demographic characteristics of cat scratch disease.

In summary, we estimate that cat scratch disease affects approximately 22 000 persons and results in the hospitalization of about 2000 persons each year in the United States. Adults appear to represent a higher proportion of cases than previously reported, suggesting that cat scratch disease may be more common in adults than formerly recognized and that it should be considered in the differential diagnosis of chronic lymphadenopathy in adults as well as children. □

Acknowledgments

We would like to thank Robert Pinner, MD, and Brian Plikaytis, MS, for their assistance.

References

1. Welch DF, Pickett DA, Slater LN, Steigerwalt AG, Brenner DJ. *Rochalimaea henselae* sp. nov., a cause of septicemia, bacillary angiomatosis, and parencymal bacillary peliosis. *J Clin Microbiol.* 1992; 30:275–280.
2. Regnery RL, Olson JG, Perkins BA, Bibb W. Serologic response of patients with suspected cat scratch disease to *"Rochalimaea henselae"* antigens. *Lancet.* 1992; 339:1443–1445.
3. Perkins BA, Swaminathan B, Jackson LA, et al. Case 22–1992—pathogenesis of cat scratch disease. *N Engl J Med.* 1992;327: 1599–1600.
4. Dolan MJ, Wong MT, Regnery RL, et al. Syndrome of *Rochalimaea henselae* adenitis suggesting cat scratch disease. *Ann Intern Med.* 1993;118:331–336.
5. Carithers HA. Cat-scratch disease: notes on its history. *Am J Dis Child.* 1970;119: 200–203.
6. Craven E, Haupt BJ. Utilization of short-stay hospitals, United States, 1980. Annual summary. *Vital Health Stat [13].* 1982; no. 64.
7. *International Classification of Diseases, 9th Revision: Clinical Modification.* Hyattsville, Md: National Center for Health Statistics; 1980. DHHS publication PHS 80-1260.
8. *U.S. Population Estimates by Age, Sex, Race, and Hispanic Origin: 1980–1991.* Washington, DC: US Bureau of the Census; 1993. Current population reports, series P-25, no. 1095.
9. Gonzalez ML, ed. *Socioeconomic Characteristics of Medical Practice 1990/1991.* Chicago, Ill: American Medical Association; 1991.
10. *Hospital Statistics 1989–90 Edition.* Chicago, Ill: American Hospital Association; 1989.
11. *Source Book of Health Insurance Data.* Washington, DC: Health Insurance Association of America; 1989.
12. Margileth AM, Wear DJ, English CK. Systemic cat scratch disease: report of 23 patients with prolonged or recurrent severe bacterial infection. *J Infect Dis.* 1987;155: 390–402.
13. Fischer GW. The agent of cat scratch disease. In: Mandell GL, Douglas RG, Bennett JE, eds. *Principles and Practice of Infectious Diseases.* 3rd ed. New York, NY: Churchill Livingstone; 1990:1874.
14. Shinall EA. Cat-scratch disease: a review of the literature. *Pediatr Dermatol.* 1990; 7:11–18.
15. Carithers HA. Cat-scratch disease: an overview based on a study of 1200 patients. *Am J Dis Child.* 1985;139:1124–1133.
16. Margileth AM. Update on cat-scratch disease. *Hosp Med.* 1989;25:61–81.
17. Margileth AM. Dermatologic manifestations and update of cat scratch disease. *Pediatr Dermatol.* 1988;5:1–9.
18. Carithers HA, Carithers CM, Edwards RO. Cat-scratch disease: its natural history. *JAMA.* 1969;207:312–316.
19. Spaulding WB, Hennessy JN. Cat scratch disease: a study of eighty-three cases. *Am J Med.* 1960;28:504–509.
20. Wlodaver CG, Vorse H. Cat-scratch disease: a view from the Oklahoma City Clinic. *J Okla State Med Assoc.* 1986;79: 679–682.
21. Moriarty RA, Margileth AM. Cat scratch disease. *Infect Dis Clin North Am.* 1987;1: 575–590.
22. Daniels WB, MacMurray FG. Cat scratch disease: report of one hundred sixty cases. *JAMA.* 1954;154:1247–1251.
23. Lewis DE, Wallace MR. Treatment of adult systemic cat scratch disease with gentamicin sulfate. *West J Med.* 1991;154:330–331.
24. Bogue CW, Wise JD, Gray GF, Edwards KM. Antibiotic therapy for cat-scratch disease? *JAMA.* 1989;262:813–816.
25. Collipp PJ. Cat-scratch disease therapy. *Am J Dis Child.* 1989;143:1261.
26. Holley HP. Successful treatment of cat-scratch disease with ciprofloxacin. *JAMA.* 1991;265:1563–1565.
27. Margileth AM. Antibiotic therapy for cat-scratch disease: clinical study of therapeutic outcome in 268 patients and a review of the literature. *Pediatr Infect Dis J.* 1992; 11:474–478.

Estimates of the US Health Impact of Influenza

ABSTRACT

Objectives. Data from the Tecumseh Community Health Study were used to estimate excess morbidity owing to influenza, and results were compared with estimates made previously using different methodology for an Institute of Medicine report.

Methods. Study participants from Tecumseh, Michigan, were classified as infected or noninfected based on laboratory results. The excess numbers of respiratory illnesses, respiratory illness days, and bed and restricted activity days experienced by the infected compared with the noninfected were estimated.

Results. The number of excess influenza-related respiratory illnesses was lower than that estimated in the Institute of Medicine report, in which all illnesses of certain characteristics occurring during an influenza season were attributed to influenza. It is now estimated that the US population under 20 years of age experiences a yearly average of 13.8 to 16.0 million influenza-related excess respiratory illnesses; for older individuals, the yearly estimate is 4.1 to 4.4 million excess illnesses.

Conclusions. For public health purposes, estimates of excess morbidity as well as of total morbidity associated with influenza should be used in setting health priorities. (*Am J Public Health.* 1993;83:1712–1716)

Kevin M. Sullivan, PhD, MPH, MHA, Arnold S. Monto, MD, and Ira M. Longini, Jr, PhD

Introduction

Influenza has been classically characterized as "the last great uncontrolled plague of mankind" because of its perceived morbidity and mortality.[1(p373)] Estimates of the morbidity associated with influenza are problematic, however, because of the difficulty of distinguishing morbidity that is influenza specific from that associated with other respiratory pathogens circulating at the same time. Thus, assumptions must be made concerning the proportion of illnesses that are influenza and the duration of influenza outbreaks during which such illnesses occur. Because of these uncertainties, the Institute of Medicine's report on *New Vaccine Development*, when attempting to prioritize the impact of various infectious diseases, stated that there were very few data available on which to estimate the "disease burden" of influenza.[2(p344)]

Over a period of 11 years, the Tecumseh Community Health Study identified influenza infection and illness among individuals of all ages in an American population.[3] This study had features that allowed the impact of influenza to be estimated based on the weekly identification of the onset of respiratory illnesses and the determination of influenza infection by virus isolation and serology, all carried out over several seasons among approximately 750 participants. Methods are now applied to data from this study to estimate the most likely amount of excess morbidity associated with influenza. Results are compared with data used in the Institute of Medicine's determination of priorities for vaccine development.[2]

Methods

Population and Study Design

Descriptions of the population and study design of the Tecumseh study have been published previously and are summarized here.[4,5] There were two periods of investigation; the current study was restricted to the years 1976 to 1981. Starting in October 1976, households in Tecumseh, Mich, were randomly selected with no restrictions concerning household eligibility; the objective was to have approximately 1000 individuals (or 10% of the population) in the study. Households were retained indefinitely, with new households being recruited to replace those that dropped out of the study. A single family respondent was contacted weekly by telephone to obtain information on respiratory or enteric illnesses in household members. The collection instrument has been described previously, as have the case definitions.[6] For each acute illness reported, details were collected on symptoms, presence of fever, and other characteristics, such as the duration of the illness and the number of restricted activity days. Illnesses were followed weekly to their conclusion. When an illness was reported within 2 days of onset, specimens for microbiological isolation were collected.[7] Blood specimens were collected regularly every 6 months.

Serologic Tests and Virus Isolation

Laboratory methods that were used have been described in detail elsewhere.[5]

Kevin M. Sullivan is with the Division of Epidemiology, and Ira M. Longini, Jr, is with the Division of Biostatistics, School of Public Health, Emory University, Atlanta, Ga. Arnold S. Monto is with the Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor.

Requests for reprints should be sent to Kevin M. Sullivan, PhD, MPH, MHA, Emory University School of Public Health, Division of Epidemiology, 1599 Clifton Rd NE, Atlanta, GA 30329.

This paper was accepted June 28, 1993.

Infections were identified by detecting a rise in hemagglutination-inhibiting antibody titer from among a series of specimens collected from an individual. Viruses used as antigen were selected for their similarity to those in circulation. Antigens were either prepared from prototype strains at the University of Michigan or obtained from the Centers for Disease Control and Prevention. Type B antigens were ether treated to increase sensitivity.[8]

For isolation of the influenza virus, specimens were inoculated into two tubes of primary monkey kidney, incubated at 33°C, and observed over 3 weeks. Hemadsorbing isolates were typed by hemagglutination inhibition.

Definitions of Outcomes and Other Terms

A participant was considered to have been infected with influenza if one or both of the following had occurred: (1) a preseason titer of less than 1:8 and a postseason titer of at least 1:16, or a fourfold rise in influenza titer during the influenza season; (2) an isolation of influenza virus. The variables studied were as follows:

- The number of respiratory illnesses. Includes the total number of respiratory illnesses, as defined by Monto et al.,[6] in the study population that had an onset during the influenza season, and the excess number of respiratory illnesses, defined as the difference between the infected and the noninfected in the mean number of respiratory illnesses that had an onset during the influenza season.
- The number of days of illness due to respiratory illness. Includes the total number of days ill with a respiratory illness in the study population that had an onset during the influenza season, and the excess number of days ill, defined as the difference between the infected and the noninfected in the mean number of days ill with a respiratory illness that had an onset during the influenza season.
- The number of bed and restricted activity days owing to respiratory illness. Includes the total number of days in which the individual was ill and stayed in bed or in which the individual's usual activities (e.g., going to work or school) were interrupted (excluding days in bed), in either case because of a respiratory illness that had an onset during the influenza season; and the excess number of days in bed or restricted in activity, defined as the difference between the infected and noninfected in the mean number of bed or restricted activity days owing to a respiratory illness that had an onset during the influenza season.

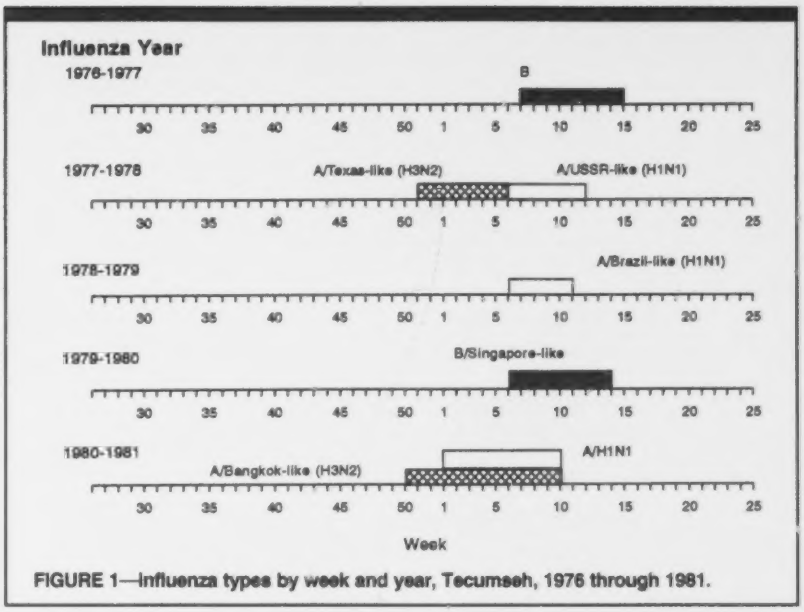


FIGURE 1—Influenza types by week and year, Tecumseh, 1976 through 1981.

The health effects of influenza were estimated by determining the excess frequency of the above health outcomes among the infected compared with the noninfected, and these results were compared with the overall frequency of the health outcomes during the influenza season. Other variables studied are as follows:

- Influenza season. The period of time in which a specific strain of influenza was isolated, during which there was an apparent excess of respiratory illnesses in the community.[5,9]
- Influenza year. The period of time in which influenza occurred, generally the last months of the calendar year and the first months of the following year. An influenza year may consist of more than one influenza season.

Age-Standardization and Sensitivity/Specificity Adjustments

All results were age adjusted using the 1980 US census data to standardize the age distribution throughout the study years. The age-standardized results were calculated for participants under 20 years of age and 20 years of age and older. The participants were divided into 5-year age groups for those under 20 years and older, but because there were few infections in certain age categories, some age groups were collapsed together. The age standardization for those under 20 years of age was based on three to four age groups, depending on the influenza season; for those 20 years of age and older, the standardization was based on two to five age groups.

Serologic tests for influenza are not 100% sensitive and specific.[1] Therefore, all results were provided with estimates adjusting for lower and higher levels of sensitivity and specificity: lower estimates were based on levels of 80% and 98%, respectively, and upper estimates were based on levels of 75% and 95%, respectively.

Results

Influenza infection occurred in Tecumseh in each of the study years. Influenza B was active in 2 years, type A(H1N1) in 3 years, and type A(H3N2) in 2 years (Figure 1). In 2 influenza years, both A(H3N2) and A(H1N1) circulated; in 1977 to 1978, the periods of circulation were separate, but in 1980 to 1981, the periods overlapped. It was thus impossible to attribute illness during the latter year to a single subtype, and so the 1980 to 1981 influenza year is omitted from the analyses of the health impact of influenza.

Table 1 shows the percentage of individuals infected in the different influenza seasons divided into the two age groups. In all years and for all influenza types and subtypes, the percentage of those infected was higher among the younger age group. The age-standardized percentage of participants infected per influenza season ranged from 5.4% to 20.4% (Table 1). For

TABLE 1—Age-Standardized[a] Frequency of Influenza Infection, by Age Group in Different Influenza Seasons, Tecumseh, 1977 through 1981

Influenza Year	Influenza Type	Age Group	n	Those Infected, %
1977	B	<20	278	21.7
		≥20	563	3.6
		Total	841	9.4
1978	A(H3N2)	<20	202	35.0
		≥20	494	13.6
		Total	696	20.4
1978	A(H1N1)	<20	183	26.5
		≥20	370	4.6
		Total	553	11.6
1979	A(H1N1)	<20	286	39.8
		≥20	458	4.1
		Total	724	15.5
1980	B	<20	286	22.0
		≥20	499	10.0
		Total	787	13.8
1981	A(H3N2)	<20	324	15.9
		≥20	490	14.3
		Total	814	14.8
1981	A(H1N1)	<20	339	14.0
		≥20	523	1.3
		Total	862	5.4

[a]Age standardized to the 1980 US population.

TABLE 2—Age-Standardized[a] Excess Number of Respiratory Illnesses Associated with Influenza and Total Number of Respiratory Illnesses per 1000 Population, Tecumseh, 1977 through 1980

Influenza Year	Influenza Type	Age Group	Excess Respiratory Illnesses/1000 Population			Total Respiratory Illnesses/ 1000 Population
			Observed	Estimated[b] Lower	Estimated[b] Upper	
1977	B	<20	81.2	114.0	126.9	793.7
		≥20	7.3	9.2	9.8	425.2
1978	A(H3N2)	<20	115.1	172.4	197.0	880.6
		≥20	14.1	18.3	19.7	520.2
1978	A(H1N1)	<20	97.6	151.6	177.7	699.4
		≥20	0.0	0.0	0.0	294.3
1979	A(H1N1)	<20	161.0	260.7	311.1	573.8
		≥20	7.2	9.1	9.8	211.8
1980	B	<20	44.5	63.4	70.9	799.6
		≥20	53.2	68.5	73.9	519.1

[a]Age standardized to 1980 US population.
[b]Lower estimate based on sensitivity and specificity levels of 80% and 98%, respectively, and upper estimate based on levels of 75% and 95%, respectively.

TABLE 3—Age-Standardized[a] Excess Number of Days with Respiratory Illness Associated with Influenza and Total Number of Days of Respiratory Illness per 1000 Population, Tecumseh, 1977 through 1980

Influenza Year	Influenza Type	Age Group	Excess Respiratory Illness Days/1000 Population			Total Respiratory Illness Days/ 1000 Population
			Observed	Estimated[b] Lower	Estimated[b] Upper	
1977	B	<20	963	1373	1526	9049
		≥20	154	195	208	5084
1978	A(H3N2)	<20	899	1326	1505	10 423
		≥20	337	436	474	6456
1978	A(H1N1)	<20	1163	1832	2156	8963
		≥20	17	22	23	3435
1979	A(H1N1)	<20	2027	3268	3892	7321
		≥20	41	52	56	2509
1980	B	<20	426	597	664	9198
		≥20	770	969	1064	6507

[a]Age standardized to 1980 US population.
[b]Lower estimate based on sensitivity and specificity levels of 80% and 98%, respectively, and upper estimate based on levels of 75% and 95%, respectively.

all age groups, there was an age-standardized average of 182 influenza infections per year (including the 1980 to 1981 influenza year) per 1000 population.

Table 2 depicts the excess number of respiratory illnesses associated with influenza infection per 1000 population. For example, in the 1977 influenza B season, there were 81.2 excess respiratory illnesses observed per 1000 population under 20 years of age. The lower and upper estimates of the "true" excess (based on different sensitivity and specificity values for the laboratory tests) are 114.0 and 126.9, respectively. With one exception, there was greater excess of respiratory illnesses per 1000 population within the younger age group compared with those aged 20 years and older. The last column presents the total number of respiratory illnesses (both influenza and noninfluenza) with an onset during the influenza season. For the younger age group, this number is, on average, 6.4 times the lower estimated number of excess respiratory illnesses and 5.6 times the upper estimate; for the older age group, the corresponding values are 26.4 and 24.6 times, respectively (excluding the 1978 A(H1N1) season).

Table 3 depicts the excess number of days of respiratory illness associated with influenza infection. Again, with the exception of 1980, the youngest age groups had a greater number of excess days of respiratory illness per 1000, and there was variability from influenza season to influenza season. The last column presents the total number of days of respiratory illnesses with an onset during the influenza season. As on Table 2, the total number of days is much greater than the excess, especially for the older age group.

The excess number of bed and restricted activity days is shown in Table 4. In each influenza season, the younger age group experienced more of such days than the older age group, and there was variability from influenza season to influenza

TABLE 4—Age-Standardized[a] Excess Number of Bed or Restricted Activity Days due to Respiratory Illnesses Associated with Influenza, and Total Number of Bed or Restricted Activity Days due to Respiratory Illnesses per 1000 Population, Tecumseh, 1977 through 1980

Influenza Year	Influenza Type	Age Group	Excess Bed or Restricted Activity Days[b]/1000 Population			Total Bed or Restricted Activity Days[b]/1000 Population
			Observed	Estimated[c] Lower	Estimated[c] Upper	
1977	B	<20	231	320	350	827
		≥20	42	53	57	254
1978	A(H3N2)	<20	362	537	611	735
		≥20	132	171	185	463
1978	A(H1N1)	<20	222	351	415	748
		≥20	0	0	0	110
1979	A(H1N1)	<20	578	968	1173	989
		≥20	14	17	18	141
1980	B	<20	296	424	474	1023
		≥20	146	188	202	367

[a]Age standardized to 1980 US population.
[b]Bed and restricted activity days due to a respiratory illness.
[c]Lower estimate based on sensitivity and specificity levels of 80% and 98%, respectively, and upper estimate based on levels of 75% and 95%, respectively.

TABLE 5—Estimates of the Impact of Influenza Extrapolated to the United States per Year, Tecumseh, 1976 through 1980

	Age Group	Estimates[a] Lower	Estimates[a] Upper
Excess respiratory illnesses, n	<20	13.8	16.0
	≥20	4.1	4.4
Excess respiratory ill days, n	<20	152.0	176.4
	≥20	66.7	70.6
Excess bed and restricted activity days, n	<20	47.1	54.7
	≥20	16.6	17.9

[a]In millions, 1980 US Census; lower estimate based on sensitivity and specificity levels of 80% and 98%, respectively, and upper estimate based on levels of 75% and 95%, respectively.

season. The last column presents the total number of bed and restricted activity days; this number is, on average, 1.9 times the lower estimated number of excess bed and restricted activity days and 1.7 times the upper estimate within the younger age group; similar values for the older age group are 4.4 and 4.2, respectively.

The estimated 4-year average excess of respiratory illnesses, respiratory ill days, and bed and restricted activity days extrapolated to the 1980 US population are shown in Table 5. All estimates are in terms of millions. The impact of influenza in the United States has been previously estimated by the Institute of Medicine based on data from the Houston Family Study[2]; these data were based on respiratory illness estimates from studies in Houston[10–12] that cover similar years as the Tecumseh data. The Tecumseh data presented here are based on excess morbidity owing to influenza, whereas the Houston data are based on a total seasonal morbidity attributed to influenza. Table 5 presents the lower and upper estimates of the excess respiratory illnesses by age group based on the Tecumseh data; because of different age categories used in the Houston data, the number of influenza-associated illnesses in one age group (15.0 to 24.9 years) was evenly divided into the two age categories used in this study. Estimates from Houston for the two age groups are approximately 23.3 million for the younger age group and 24.4 million for the older age group (based on extrapolation of the Houston experience to the 1984 US population).

Discussion

To define influenza, this analysis used the excess number of total respiratory illnesses rather than a constellation of symptoms because the symptoms associated with influenza differ by age.[5] In all influenza seasons, individuals infected with the influenza virus had more respiratory illnesses, more days of respiratory illness, and more bed and restricted activity days than those not infected, but the excess was not homogeneous across the age groups.[13] The duration of the influenza seasons differed, and the selection of the beginning and ending of an influenza season was somewhat arbitrary. However, use of the difference (or excess) between infected and noninfected individuals should not be affected by the duration of the outbreak as long as the defined season covers the time of influenza activity.

In Tecumseh, use of the total number of health events (number of respiratory illnesses, days of respiratory illness, and bed and restricted activity days) would overestimate the excess number of health events associated with influenza infection, especially for the older age group. The more severe the outcome (from the number of days with respiratory illness to the number of severely ill [i.e., bed and restricted activity] days), the less the total number of events overestimated the excess number of events. It may be that, for the most severe health effects of influenza (such as hospitalization or mortality), the total number of events will overestimate the excess number even less, although this requires further study.

Reasons for the differences between the Tecumseh and Houston estimates appear mainly to relate to the different methods used to measure the outcomes. The Tecumseh data are based on excess respiratory illnesses whereas the Houston data are based on all respiratory illnesses occurring during an influenza season. It is possible that attributing all respiratory illnesses that occur during influenza seasons to influenza overestimates the disease burden of influenza. Another possible reason for the differences in these estimates could be presented by the following equation:

Total = Excess + Replacement,
where Total is the number of influenza illnesses,
Excess is the frequency of illnesses in the infected above the frequency in the noninfected, and
Replacement is when an influenza illness replaces a respiratory illness in those infected that would have been caused by another agent had influenza not occurred.

If influenza replaces other infectious agents among the infected during influenza outbreaks, then the total number of respiratory illnesses caused by influenza

would be greater than the excess alone. The excess number of respiratory illnesses caused by influenza provides a minimum estimate of the overall disease burden of influenza. Influenza infection may replace an infection by another respiratory agent,[5] or, if influenza infection is prevented, it may be replaced by another respiratory agent. From a cost-benefit point of view, measurement of the excess amount of the disease burden would be more valuable.

An unexpected finding was the severity of influenza B in the 1980 season among the adults. In terms of excess respiratory illnesses and excess days with a respiratory illness, this was the most severe season studied among the adults. These results are consistent, however, with the severity of this type B outbreak, which produced excess pneumonia and influenza deaths nationwide.[14]

The results presented are consistent with previous studies showing that influenza results in excess morbidity. One possible problem in these analyses is the potential for "cross-protective" effects between sequential A(H3N2) and A(H1N1) outbreaks in the same season.[15–17] Another is the sensitivity and specificity of the laboratory tests used, which would make the observed excess morbidity biased toward the null value of no difference between infected and noninfected individuals. With lower sensitivity and specificity estimates of 80% and 98%, respectively, and higher estimates of 75% and 95% respectively, the true health effects would be greater than the observed excess. The sensitivity and specificity of influenza laboratory tests most likely vary by influenza strain, age, and perhaps other factors. In this study, the proportion infected in each age group was treated as a fixed value. The estimates of the proportion infected is a random variable and is subject to sampling variability, which was not taken into account in these analyses.

The interpretation of the Tecumseh data extrapolated to the United States must be done with caution; this extrapolation was carried out principally to facilitate comparison with results presented in the Institute of Medicine's report, which used data derived from another community.[2] First, the data are based on fewer than 1000 participants per influenza season. Second, Tecumseh is located in a rural area in southeast Michigan, and it is not clear if the epidemiology of influenza in this community reflects the average experience in the United States. Third, it is difficult to compare estimates of total vs excess influenza morbidity. However, from a public health viewpoint, the latter would seem to be of greater importance. Thus, the estimates of the disease burden of influenza provided here may be useful for cost-benefit studies and policy purposes. □

Acknowledgments

We thank Drs Hunein F. Maassab and Lisa A. Weisfeld for their comments, and the Centers for Disease Control and Prevention for influenza antigens.

References

1. Davenport FM. Influenza viruses. In: Evans AS, ed. *Viral Infections of Humans: Epidemiology and Control.* 2nd ed. New York, NY: Plenum Medical Book Co; 1984:373–397.
2. Committee on Issues and Priorities for New Vaccine Development, Institute of Medicine. *New Vaccine Development: Establishing Priorities.* Vol 1. *Diseases of Importance in the United States.* Washington, DC: National Academy Press; 1985; 342–364.
3. Monto AS, Cavallaro JJ. The Tecumseh study of respiratory illness: II. patterns of occurrence of infection with respiratory pathogens, 1965–1969. *Am J Epidemiol.* 1971;94:280–289.
4. Monto AS, Lim SK. The Tecumseh study of respiratory illness: III. incidence and periodicity of respiratory syncytial virus and *Mycoplasma pneumoniae* infections. *Am J Epidemiol.* 1971;94:290–301.
5. Monto AS, Koopman JS, Longini IM Jr. Tecumseh study of illness: XII. influenza infection and disease, 1976–1981. *Am J Epidemiol.* 1985;121:811–822.
6. Monto AS, Napier JA, Metzner HL. The Tecumseh study of respiratory illness: I. Plan of study and observation on syndromes of acute respiratory disease. *Am J Epidemiol.* 1971;94:269–279.
7. Monto AS, Kiomehr F. The Tecumseh study of respiratory illness: IV. occurrence of influenza in the community, 1966–1971. *Am J Epidemiol.* 1975;102:553–563.
8. Monto AS, Maassab HF. Ether treatment of type B influenza virus antigen for the hemagglutination inhibition test. *J Clin Microbiol.* 1981;13:54–57.
9. Longini IM, Koopman JS, Haber M, et al. Statistical inference for infectious diseases: risk-specific household and community transmission parameters. *Am J Epidemiol.* 1988;128:845–859.
10. Glezen WP. Serious morbidity and mortality associated with influenza epidemics. *Epidemiol Rev.* 1982;4:25–44.
11. Glezen WP, Frank AL, Taber LH, et al. Influenza in childhood. *Pediatr Res.* 1984; 17:1029–1032.
12. Glezen WP, Six HR, Perrotta DM, et al. Epidemics and their causative viruses—community experience. In: Stuart-Harris C, Potter CW, eds. *The Molecular Virology and Epidemiology of Influenza.* New York, NY: Academic Press; 1984.
13. Sullivan KM. *Influenza Infection, Pathogenicity, and Virulence: Tecumseh, 1976–1981.* Ann Arbor, Mich: University of Michigan; 1990. PhD diss.
14. Lui KJ, Kendal AP. Impact of influenza epidemics on mortality in the United States from October 1972 to May 1985. *Am J Public Health.* 1987;77:712–716.
15. Francis T Jr, Davenport FM, Hennessy AV. A serological recapitulation of human infection with different strains of influenza virus. *Trans Assoc Am Physicians.* 1953; 66:231–239.
16. Sonoguchi T, Naito H, Hara M, et al. Cross-subtype protection in humans during sequential, overlapping, and/or concurrent epidemics caused by H3N2 and H1N1 influenza viruses. *J Infect Dis.* 1985;151:81–99.
17. Ackerman E, Longini IM Jr, Seaholm SK, et al. Simulation of mechanisms of viral interference in influenza. *Int J Epidemiol.* 1990;2:444–454.

The Risk of Measles, Mumps, and Varicella among Young Adults: A Serosurvey of US Navy and Marine Corps Recruits

ABSTRACT

Objectives. To assess the risk of epidemic transmission and to guide immunization policy, the seroprevalence of antibody to measles, mumps, and varicella was determined in a group of young adults.

Methods. A cross-sectional study of 1533 US Navy and Marine Corps recruits was conducted in June 1989. Antibody status was determined with commercially available enzyme-linked immunosorbent assays.

Results. Direct sex and race adjustment to the 15- to 29-year-old US population resulted in seronegativity rates of 17.8% for measles, 12.3% for mumps, and 6.7% for varicella. Measles and mumps seronegativity rates were higher among Whites whereas varicella seronegativity was higher among non-Whites. Recruits enlisting from outside the 50 US states, especially those from island territories, were more likely to lack varicella antibody. The sensitivity of a positive history of vaccination or disease in predicting antibody status was less than 90% for all diseases.

Conclusions. These results suggest a continued potential for epidemics, especially of measles, and the need for mandatory immunization policies. Immigrants to the United States, especially those from island territories, may be a high-risk group that could benefit from varicella vaccination. (*Am J Public Health.* 1993; 83:1717–1720)

Jeffery P. Struewing, MD, MS, Kenneth C. Hyams, MD, MPH, John E. Tueller, MD, MPH, and Gregory C. Gray, MD, MPH

Introduction

Despite great progress in the United States over the last several decades in controlling vaccine-preventable illnesses,[1,2] recent increases in the incidence of measles and mumps—including epidemics—have occurred,[3,4] especially among inner-city children and young adult populations.[5–13] These viral illnesses also continue to have considerable military importance.[14–18] In addition, varicella, a serious illness in adults,[19–21] may also be increasing in incidence, particularly in military populations.[22]

Recruits enter the service from all areas of the country and most are between 17 and 19 years old, an age group at high risk in the civilian setting for measles, mumps, and varicella. Because these viral illnesses may have a great impact on health, training time, and overall military readiness, especially in the recruit setting, and because their importance is increasing in the civilian population, a serosurvey was conducted to guide immunization policy. In addition, demographic factors and other information related to individuals' previous disease and vaccination history were evaluated as potential predictors of antibody status.

Methods

This study was conducted in June 1989 at the three US Navy (Great Lakes, Ill; Orlando, Fla; San Diego, Calif) and the two US Marine Corps (Parris Island, SC; San Diego, Calif) recruit training centers. Four consecutive companies/platoons at each location were enrolled in the study. Those recruits who gave informed consent completed a short questionnaire and donated an extra tube of blood.

Blood samples were obtained 1 to 4 days after the recruits' arrival at the training center and before the administration of immunizations. Serum samples were tested for measles, mumps, and varicella antibodies using commercially available enzyme-linked immunosorbent assay (ELISA) tests (MEASLESTAT, MUMPSTAT, and VARICELISA II from Whittaker Bioproducts, Walkersville, Md). To assess interrater reliability, 150 randomly selected samples were retested by the manufacturer. The kappa value for varicella retesting was 0.96, and no further retesting was performed. Because the kappa values for measles and mumps were lower (0.83 and 0.77, respectively), the manufacturer retested all samples for these two antibodies in the laboratory, and the manufacturer's results were used in the analysis. Samples were initially tested for rubella antibody as well, but extremely high seronegativity rates suggested technical difficulties with the procedure, and insufficient serum was available to allow retesting of samples. Therefore, no rubella results are presented. All testing was done without knowledge of the subjects' prior disease or vaccination history.

The self-completed questionnaire included demographic variables and information on state of birth, elementary

Jeffery P. Struewing, John E. Tueller, and Gregory C. Gray are with the Navy Environmental and Preventive Medicine Unit No. 5, San Diego, Calif. Kenneth C. Hyams is with the Epidemiology Division of the Naval Medical Research Institute, Bethesda, Md.

Requests for reprints should be sent to Jeffrey P. Struewing, MD, MS, National Institutes of Health, Genetic Epidemiology Branch, Bldg EPN-439, Bethesda, MD 20892-9903.

This paper was accepted June 23, 1993.

Note. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, or the US government.

TABLE 1—Sample's Seronegativity to Measles, Mumps, and Varicella, by Demographic Variables

	Men				Women			
	No. Tested	% Seronegative (95% Confidence Intervals)			No. Tested	% Seronegative (95% Confidence Intervals)		
		Measles	Mumps	Varicella		Measles	Mumps	Varicella
Total sample	1392	19.2 (17.1, 21.3)	12.2 (10.5, 13.9)	9.2 (7.7, 10.7)	141	14.2 (8.4, 20.0)	7.1 (2.9, 11.3)	7.1 (2.9, 11.3)
Enlistees from 50 US states	1303	19.5 (17.3, 21.7)	12.2 (10.4, 14.0)	8.3 (6.8, 9.8)	139	14.4 (8.6, 20.2)	7.2 (2.9, 11.5)	7.2 (2.9, 11.5)
Age group, y								
17–19	983	20.3 (17.8, 22.8)	13.2 (11.1, 15.3)	8.2 (6.5, 9.9)	86	15.1 (7.5, 22.7)	5.8 (0.9, 10.7)	4.7 (0.2, 9.2)
20–24	269	16.7 (12.2, 21.2)	9.7 (16.2, 13.2)	8.2 (4.9, 11.5)	40	12.5 (2.3, 22.7)	5.0 (0, 11.8)	15.0 (3.9, 26.1)
25+	42	14.3 (3.7, 24.9)	2.4 (0.0, 7.0)	11.9 (2.1, 21.7)	12	16.7 (0, 37.8)	25.0 (0.5, 49.5)	0.0 (0, 22.1)
Education								
< High school	74	24.3 (14.5, 34.1)	20.3 (11.1, 29.5)	13.5 (5.7, 21.3)	0	...	...	...
High school graduate	976	18.9 (16.4, 21.4)	13.0 (10.9, 15.1)	7.3 (5.7, 8.9)	96	16.3 (9.0, 23.6)	7.1 (2.0, 12.2)	7.1 (2.0, 12.2)
> High school graduate	251	20.7 (15.7, 25.7)	6.8 (3.7, 9.9)	10.8 (7.0, 14.6)	41	9.8 (0.7, 18.9)	7.3 (0, 15.3)	7.3 (0, 15.3)
Race								
White	926	22.4 (19.7, 25.1)	13.8 (11.6, 16.0)	6.5 (4.9, 8.1)	87	17.2 (9.3, 25.1)	10.3 (3.9, 16.7)	4.6 (0.2, 9.0)
Black	221	9.5 (5.6, 13.4)	8.1 (4.5, 11.7)	13.6 (9.1, 18.1)	32	15.6 (3.0, 28.2)	3.1 (0, 9.1)	9.4 (0, 19.5)
Hispanic	121	19.0 (12.0, 26.0)	8.3 (3.4, 13.2)	10.7 (5.2, 16.2)	12	0.0 (0, 22.1)	0.0 (0, 22.1)	25.0 (0.5, 49.5)
Other	33	6.1 (0, 14.3)	9.1 (0, 18.9)	15.2 (3.0, 27.4)	8	0.0 (0, 31.2)	0.0 (0, 31.2)	0.0 (0, 31.2)

Note. Age was missing for 9 males and 1 female, education for 2 males, and area of enlistment for 5 subjects.

school entry, and military enlistment. Subjects were asked to recall their history of having had each of the individual diseases and vaccinations, with four possible responses: (1) yes, definitely; (2) yes, probably; (3) no, definitely not; and (4) I don't know.

The Centers for Disease Control and Prevention reports of school entry immunization laws for each state were available for 1970 to 1977, 1979 to 1980, 1983, and 1985 to 1990. These reports were used to determine the year-specific state school entry immunization laws for each recruit.

Statistical analyses were conducted with EGRET statistical software (SERC, Inc, Seattle, Wash) on a microcomputer. To compare the proportion of recruits who were seronegative, logistic regression for the continuous variable of age, the Mantel-Haenszel chi-square test for trend in education, and the likelihood ratio chi-square for other variables were used. Adjustment for potential confounders in the association between race and seronegativity was performed using logistic regression models.

The proportion of seropositive recruits who reported a positive history of disease or vaccination was defined as the sensitivity of the historical information, and the proportion of seronegative recruits who did not report prior disease or vaccination was defined as the specificity. Positive predictive value was defined as the proportion of history-positive recruits who were seropositive, and the negative predictive value was defined as the proportion of history-negative recruits who were seronegative.

Results

Of the 1568 recruits assigned to the selected companies, 1533 (98%) gave informed consent and participated in the study. Among the 1442 subjects enlisting from one of 48 US states, 90% were male, 74% were aged 17 to 19, 70% were White, 18% were Black, and 9% were Hispanic. Seronegativity rates and 95% confidence intervals were as follows: measles, 19.0% (17.0, 21.0); mumps, 11.7% (9.6, 13.8); and varicella, 8.2% (6.8, 9.6). Direct sex and race adjustment to the 1980 US 15- to 29-year-old population resulted in estimated national seronegativity rates of 17.8% for measles, 12.3% for mumps, and 6.7% for varicella. Seronegativity to measles and mumps was not highly correlated, with only 2.8% seronegative for both.

Men were more likely to be seronegative for all three viruses, and results are presented separately for men and women (Table 1). Younger male recruits were more likely than older ones to be seronegative for measles and mumps ($P > .05$), but older female recruits were more likely than younger ones to be seronegative for mumps ($P < .05$). Age was not associated with varicella antibody status for either men or women. Higher educational level was associated only with lower mumps seronegativity in men ($P < .01$). Among both men and women, Whites were more likely to be seronegative for measles and mumps while non-Whites were much more likely to be seronegative for varicella antibody ($P < .05$ for men). The association between race and antibody status for all three viruses persisted after controlling for age, sex, and educational level.

Recruits in this study had enlisted from 48 of the 50 states, and although there was wide variation in seronegativity between states, no clear geographic trends were evident. Recruits whose state of enlistment had full mumps immunization entry laws (kindergarten through grade 12) were slightly less likely to be seronegative for mumps (9.8%) compared with other subjects (13%). No other associations between state immunization laws and measles or mumps antibody status were present, even when those recruits who reported the same state of birth, elementary school entry, and military enlistment were analyzed separately.

Eighty-four recruits (5%) enlisted from outside the United States. Except for those recruits enlisting directly from the Republic of the Philippines, most of whom receive measles/rubella immunization as part of their processing before arrival in the United States, their rates of measles

and mumps seronegativity were similar to those of US enlistees. However, more than 20% of recruits enlisting from outside the United States, including 3 of 3 from the Canal Zone, 2 of 3 from Puerto Rico, and 6 of 23 enlisting from the Republic of the Philippines, were varicella seronegative.

Combining recruits who reported a positive history of disease *or* vaccination, the sensitivity of a "definite" history in identifying seropositive recruits was less than 35% for both measles and mumps (Table 2), whereas the sensitivity of a definite history of varicella disease was 75%. Definite *or* "probable" histories of disease or vaccination doubled the sensitivity for measles and mumps, and increased the sensitivity of varicella disease history to 86%. Corresponding specificities were similar, ranging from 35% to 72% for measles and mumps, to 66% to 75% for varicella. The sensitivity of vaccine history was generally better than the disease history.

The negative predictive value was lower for mumps than for measles or varicella. On the other hand, the positive predictive value of a history of varicella was over 95%. However, almost all recruits gave a positive history of varicella, including more than 20% of those lacking the varicella antibody.

TABLE 2—The Ability of Undocumented History of Disease or Vaccination to Predict Serum Antibody Status to Measles, Mumps, and Varicella

Self-Reported Historical Information	No. with Positive History/Total	Sensitivity[a]	Specificity[b]	PV +[c]	PV –[d]
Definitely had measles disease or vaccine	470/1434	33.5	70.4	82.8	20.0
Definitely or probably had measles disease or vaccine	935/1434	66.2	35.0	81.0	19.2
Definitely had mumps disease or vaccine	472/1436	33.5	72.1	90.0	12.7
Definitely or probably had mumps disease or vaccine	959/1436	67.5	38.5	89.2	13.6
Definitely had varicella disease	421/1437	74.8	75.2	97.1	20.9
Definitely or probably had varicella disease	1171/1437	85.7	65.8	98.6	28.9

Note. Individual disease/vaccine totals fall below sample total because of missing values.
[a]Sensitivity is the proportion of antibody-positive subjects who gave a positive history.
[b]Specificity is the proportion of antibody-negative subjects who gave a negative history.
[c]PV+ is the positive predictive value, or the proportion of history-positive subjects who were antibody positive.
[d]PV– is the negative predictive value, or the proportion of history-negative subjects who were antibody negative.

Discussion

Based on commercially available ELISA tests, nearly one fifth of the 1533 recruits in this study lacked antibody to measles, 12% lacked antibody to mumps, and 9% were seronegative for varicella antibody. Direct adjustment to the 1980 US 15- to 29-year-old population yielded similar results. If seronegativity correlates with susceptibility, these results suggest the potential for continued epidemic transmission of measles in young adult populations. However, a routine two-dose schedule of the measles/mumps/rubella vaccination, as has been recommended,[23,24] should considerably reduce this potential.

The direct US adjusted measles seronegativity rate (17.8%) observed in this study was higher than expected, based on several serosurveys in the past 2 decades,[15,25,26] but was comparable to the rate of 20.7% in a recent study of US Army recruits.[27] Fewer previous mumps serosurveys have been reported,[28] but the adjusted mumps seronegativity rate of 12.3% in this study was comparable to that found in the most recent US Army recruit survey.[27] The 6.7% adjusted varicella seronegativity rate was similar to that found in two recent studies[27,29] but lower than that found in another.[26]

The lower seronegativity rate for measles and mumps among older male subjects in this study probably reflects increased natural infection during childhood. Among the group born before 1964, vaccination was not widespread when most subjects entered elementary school. The lack of association between age and varicella seronegativity rates observed in this study suggests no decrease in natural infection rates in recent years.

Whites were more likely to be seronegative to both measles and mumps and less likely to be seronegative to varicella, even after adjustment for age, sex, and education. This finding may reflect lower rates of natural measles and mumps infection among Whites or higher rates of waning vaccine-induced antibody levels. The associations between race and varicella seronegativity and between sex and seronegativity to all three viruses may reflect unknown sociocultural differences in patterns of natural infection or vaccination. Recruits enlisting from outside the 50 US states, especially those from island nations or territories, were much more likely to be seronegative for varicella antibody. This supports the previous epidemiological observations of an island phenomenon.[21,30] Especially in the health care and military recruit environments, foreign-born individuals may represent a high-risk group that could benefit from a varicella vaccine.[31]

The specificity of historical information of measles disease or vaccination was higher in this study than in a prior study of medical students.[26] However, using this historical information to determine which recruits should be vaccinated or receive antibody tests would not effectively reduce the proportion remaining seronegative to below 8%. Both the sensitivity and specificity of a history of varicella in this study were generally higher than in a previous study.[26] Most recruits, however, reported a positive history of disease, and more than 20% of seronegative recruits would be missed if such a screening question were used to determine who would receive varicella immunization. The lower negative predictive value for mumps compared with that for measles and varicella is probably owing to the fact that mumps infection is often subclinical.

Several factors need to be considered when interpreting the results of this serosurvey. Although the ELISA tests used here appear to be as good as previous antibody tests when compared to a gold standard,[32–38] they may not correlate with true susceptibility to these illnesses. However, given the evidence for the role of waning vaccine-induced antibody levels in recent outbreaks of measles,[39,40] a less sensitive test may be desirable if it is being used to decide who gets immunized in the recruit setting. Although the true number of young adults currently susceptible to these illnesses may be considerably lower than this serosurvey suggests, these results may be predictive of the future potential for epidemics, especially if the intensity of exposure increases, because

vaccine-induced antibody titers may continue to wane.

Subjects in this study came from most areas of the United States, but whether they are representative of the civilian population with respect to their infection and vaccination experiences and associated risk factors is unknown. This was a cross-sectional study done over a short period, and further studies of these subjects and cross-sectional studies over time will be important in monitoring seroprevalence trends.

The high rates of seronegativity among US recruits suggest the continued potential for epidemics of these illnesses and the need for continued effective immunization programs in both the civilian and military setting. In the recruit training center environment, historical information on previous disease and vaccination history would not be an effective guide in determining who should be immunized upon entry into the US military. □

Acknowledgments

Financial support for this work was provided by the Naval Medical Research and Development Command, National Naval Medical Command, National Capitol Region, Bethesda, Md, Work Unit no. 3M162770A870AR1.

The authors wish to thank the medical staff of the Naval Training Center Great Lakes, Naval Training Center Orlando, Naval Training Center San Diego, Marine Corps Recruit Depot Parris Island, Marine Corps Recruit Depot San Diego, Naval Hospital Beaufort, Naval Hospital Great Lakes, Naval Hospital Orlando, and Naval Hospital San Diego; and Dr Owen Wood and Stephanie Gray, Naval Medical Research Institute, Bethesda, Md.

References

1. Centers for Disease Control. Measles—United States, 1988. *MMWR.* 1989;38:601–605.
2. Centers for Disease Control. Mumps—United States, 1985–1988. *MMWR.* 1989; 38:101–105.
3. Centers for Disease Control. Measles—United States, first 26 weeks, 1989. *MMWR.* 1989;38:863–866, 871–872.
4. Cochi SL, Preblud SR, Orenstein WA. Perspectives on the relative resurgence of mumps in the United States. *Am J Dis Child.* 1988;142:499–507.
5. Markowitz LE, Preblud SR, Orenstein WA, et al. Patterns of transmission in measles outbreaks in the United States, 1985–1986. *N Engl J Med.* 1989;320:75–81.
6. Centers for Disease Control. Measles—Los Angeles County, California, 1988. *MMWR.* 1989;38:49–52, 57.
7. Krause PJ, Cherry JD, Deseda-Tous J, et al. Epidemic measles in young adults: clinical, epidemiological, and serologic studies. *Ann Intern Med.* 1979;90:873–876.
8. Centers for Disease Control. Multiple measles outbreaks on college campuses—Ohio, Massachusetts, Illinois. *MMWR.* 1985;34:129–130.
9. Centers for Disease Control. Measles on college campuses—United States, 1985. *MMWR.* 1985;34:445–448.
10. Centers for Disease Control. Mumps outbreaks on university campuses—Illinois, Wisconsin, South Dakota. *MMWR.* 1987; 36:496–498, 503–505.
11. Wharton M, Cochi SL, Hutcheson RH, Bistowish JM, Schaffner W. A large outbreak of mumps in the postvaccine era. *J Infect Dis.* 1988;158:1253–1260.
12. Sosin DM, Cochi SL, Gunn RA, Jennings CE, Preblud SR. Changing epidemiology of mumps and its impact on university campuses. *Pediatrics.* 1989;84:779–784.
13. Kaplan KM, Marder DC, Cochi SL, Preblud SR. Mumps in the workplace. Further evidence of the changing epidemiology of a childhood vaccine-preventable disease. *JAMA.* 1988;260:1434–1438.
14. Black FL. A nationwide serum survey of United States military recruits, 1962: III. measles and mumps antibodies. *Am J Hyg.* 1964;80:304–307.
15. Crawford GE, Gremillion DH. Epidemic measles and rubella in Air Force recruits: impact of immunization. *J Infect Dis.* 1981; 144:403–410.
16. Liao SJ. Immunity status of military recruits in 1951 in the United States: II. results of mumps complement-fixation tests. *Am J Hyg.* 1954;59:273–281.
17. Longfield JN, Winn RE, Gibson RL, Juchau SV, Hoffman PV. Varicella outbreaks in Army recruits from Puerto Rico: varicella susceptibility in a population from the tropics. *Arch Intern Med.* 1990;150: 970–973.
18. Gray GC, Struewing JP, Clever LG, Kuhlenkamp JF. Worst California measles epidemic in 25 years. *Navy Med.* 1991;82:23–25.
19. Preblud SR. Varicella: complications and costs. *Pediatrics.* 1986;78(suppl):728–735.
20. Guess HA, Broughton DD, Melton LJ, Kurland LT. Population-based studies of varicella complications. *Pediatrics.* 1986; 78(suppl):723–727.
21. Weller TH. Varicella-herpes zoster virus. In: Evans AS, ed. *Viral Infections of Humans: Epidemiology and Control.* New York, NY: Plenum Medical Book Co; 1984:666.
22. Gray GC, Palinkas LA, Kelley PW. Increasing incidence of varicella hospitalizations in United States Army and Navy personnel: are today's teenagers more susceptible? Should recruits be vaccinated? *Pediatrics.* 1990;86:867–873.
23. Measles prevention: recommendations of the Immunization Practices Advisory Committee (ACIP). *MMWR.* 1989;38:1–18.
24. American Academy of Pediatrics Committee on Infectious Diseases. Measles: reassessment of the current immunization policy. *Pediatrics.* 1989;84:1110–1113.
25. Preblud SR, Gross F, Halsey NA, Hinman AR, Herrmann KL, Koplan JP. Assessment of susceptibility to measles and rubella. *JAMA.* 1982;247:1134–1137.
26. Murray DL, Lynch MA. Determination of immune status to measles, rubella, and varicella-zoster viruses among medical students: assessment of historical information. *Am J Public Health.* 1988;78:836–838.
27. Kelley PW, Petruccelli BP, Stehr-Green P, Erickson RL, Mason CJ. The susceptibility of young adult Americans to vaccine-preventable infections: a national serosurvey of US Army recruits. *JAMA.* 1991;266: 2724–2729.
28. Shehab ZM, Brunell PA, Cobb E. Epidemiological standardization of a test for susceptibility to mumps. *J Infect Dis.* 1984; 149:810–812.
29. McKinney WP, Horowitz MM, Battiola RJ. Susceptibility of hospital-based health care personnel to varicella-zoster virus infections. *Am J Infect Control.* 1989;17:26–30.
30. Maretic Z, Cooray MPM. Comparisons between chickenpox in a tropical and a European country. *J Trop Med Hyg.* 1963;66: 311–315.
31. Nassar NT, Touma HC. Susceptibility of Filipino nurses to the varicella-zoster virus. *Infect Control.* 1986;7:71–72.
32. Boteler WL, Luipersbeck PM, Fuccillo DA, O'Beirne AJ. Enzyme-linked immunosorbent assay for detection of measles antibody. *J Clin Microbiol.* 1983;17:814–818.
33. Glikmann G, Mordhorst CH. Serological diagnosis of mumps and parainfluenza type-1 virus infections by enzyme immunoassay, with a comparison of two different approaches for detection of mumps IgG antibodies. *Acta Pathol Microbiol Immunol Scand.* 1986;94:157–166.
34. Nigro G, Nanni F, Midulla M. Determination of vaccine-induced and naturally acquired class-specific mumps antibodies by two indirect enzyme-linked immunosorbent assays. *J Virol Methods.* 1986;13:91–106.
35. Larussa P, Steinberg S, Waithe E, Hanna B, Holzman R. Comparison of five assays for antibody to varicella-zoster virus and the fluorescent-antibody-to-membrane-antigen test. *J Clin Microbiol.* 1987;25:2059–2062.
36. Neumann PW, Weber JM, Jessamine AG, O'Shaughnessy MV. Comparison of measles antihemolysin test, enzyme-linked immunosorbent assay, and hemagglutination inhibition test with neutralization test for determination of immune status. *J Clin Microbiol.* 1985;22:296–298.
37. Cremer NE, Cossen CK, Sheil G, Diggs J, Gallo D, Schmidt NJ. Enzyme immunoassay versus plaque neutralization and other methods for determination of immune status to measles and varicella-zoster and versus complement fixation serodiagnosis of infections with those viruses. *J Clin Microbiol.* 1985;21:869–874.
38. Sakata H, Hishiyama M, Sugiura A. Enzyme-linked immunosorbent assay compared with neutralization tests for evaluation of live mumps vaccines. *J Clin Microbiol.* 1984;19:21–25.
39. Wassilak SG, Orenstein WA, Strickland PL, Butler CA, Bart KJ. Continuing measles transmission in students despite a school-based outbreak control program. *Am J Epidemiol.* 1985;122:208–217.
40. Mathias RG, Meekison WG, Arcand TA, Schechter MT. The role of secondary vaccine failures in measles outbreaks. *Am J Public Health.* 1989;79:475–478.

Multiethnic Variations in the Pregnancy Outcomes of Military Dependents

ABSTRACT

Objectives. We examined the pregnancy outcomes of three ethnic groups: African-American Blacks, non-Hispanic Whites, and Filipinos. In an attempt to reduce ethnic dissimilarities in parental employment and access to health care, this investigation compared the single-live-birth outcomes of married, adult women who resided in the state of Hawaii and who indicated that their spouse was on active-duty status in the US military.

Methods. The data for this study were obtained from the 1979–1989 Hawaii vital-record file that provides linked live birth–infant death information. Multiple logistic regression was used to calculate odds ratios for the independent effects of maternal factors on low birthweight and neonatal mortality.

Results. Significant differences in maternal age, maternal education, paternal education, parity, hospital of delivery, and use of prenatal care were observed among the ethnic groups. The results of a logistic regression analysis of low birthweight indicated significantly higher risks for Filipinos and Blacks compared with Whites. For very low birthweight, only an increased risk for Blacks was observed. No ethnic differences in neonatal mortality were found.

Conclusions. This investigation revealed more comparable infant mortality experiences among the ethnic groups in spite of persistent birthweight differences. (*Am J Public Health.* 1993;83:1721–1725)

Greg R. Alexander, ScD, MPH, Gigliola Baruffi, MD, MPH, Joanne M. Mor, MS, Edith C. Kieffer, PhD, MPH, and Thomas C. Hulsey, ScD, MSPH

Introduction

Although infant mortality rates have declined markedly during the last century in the United States, ethnic disparities in infant mortality and related measures of pregnancy outcome persist; with respect to some ethnic comparisons, notably between White and Black, these disparities have increased.[1] The factors that underlie these ethnic variations have become an important area of research in the public health, social science, and medical fields. However, in spite of ongoing programmatic and policy efforts to close the gap in measures of pregnancy outcome among ethnic groups, there is reason for serious concern that these disparities will extend to another generation of infants.

Numerous studies have revealed that these ethnic variations in pregnancy outcome can be partially explained by socioeconomic factors, demographic factors, extent of prenatal-care utilization, and conditions associated with medical risk.[1–19] However, many of the observed ethnic differences in pregnancy outcomes remain unexplained. Controlling for disparities in employment status, living conditions, and access to and use of health care resources represents a formidable challenge to studies of ethnic variation in pregnancy outcomes. In investigations using data from vital records, maternal education, age, and marital status are typically the only variables available to indicate socioeconomic status and to adjust for disparities in socioeconomic status among ethnic populations. Although indices of prenatal-care utilization have been used to control for ethnic variations in adequacy of prenatal care, these indices do not consider the influences of differences in the content and quality of prenatal care received by various ethnic groups.[8,20,21] Finally, much of the US scientific literature on ethnic differences in pregnancy outcomes has been centered on the comparison of Black and White populations. Only a few studies have considered multiple ethnic groups or have further attempted to control for content and quality of prenatal care.[5,22] Notwithstanding the appreciable and disturbing contrasts provided by the prevailing research focus on Black–White biethnic differences, the inclusion of other ethnic groups in future investigations may offer new insights.

In this study, we examine the pregnancy outcomes of three ethnic groups: African-American Blacks, non-Hispanic Whites, and Filipinos. In an attempt to reduce ethnic dissimilarities in parental employment, general living environment, and access to health care, this investigation compares the single-live-birth outcomes of married, adult women who resided in the state of Hawaii and who indicated that the father of the infant had an active-duty US military status. Previous comparisons of ethnic variation in pregnancy outcome among military populations have been largely limited to Black and White populations.[15,18,19,23]

The investigation of data from the state of Hawaii offers several unique ad-

Greg R. Alexander, Gigliola Baruffi, Joanne M. Mor, and Edith C. Kieffer are with the School of Public Health, University of Hawaii at Manoa. Greg R. Alexander is also with the School of Public Health, University of Minnesota, Minneapolis. Thomas C. Hulsey is with the Department of Pediatrics, Medical University of South Carolina, Charleston.

Requests for reprints should be sent to Greg R. Alexander, ScD, MPH, Maternal and Child Health Major, School of Public Health, University of Minnesota, Box 97 UHMC, 420 Delaware St SE, Minneapolis, MN 55455.

This paper was accepted April 26, 1993.

TABLE 1—Parental and Prenatal-Care Characteristics, by Ethnicity of Mother

Characteristic	Ethnicity of Mother, %		
	White (n = 24 674)	Filipino (n = 4521)	Black (n = 4617)
≥35 years of age	3.04	7.10	1.84
<12 years of education (mother)	8.20	17.02	6.29
Primipara	46.29	37.56	37.58
High parity for age[a]	0.52	0.77	0.82
>12 years of education (father)	39.68	41.77	28.95
Military hospital delivery	82.46	78.50	86.51
First-trimester prenatal care	81.58	75.38	77.39
Inadequate or no prenatal care[b]	3.23	5.82	5.24

Note. Data are for single live births to adult, married, military-dependent mothers residing in Hawaii 1979 through 1989. *P* value (chi-square test) for all characteristics is $P < .01$.
[a]High parity for age: for women 18–21 years, three or more births; for women 22–24 years, four or more births; for women 25+ years, five or more births.
[b]Institute of Medicine Index of Adequacy of Prenatal Care.[20]

TABLE 2—Pregnancy Outcomes, by Ethnicity of Mother

Outcome	Ethnicity of Mother			
	White	Filipino	Black	*P*
Very preterm,[a] %	1.10	1.59	2.62	<.01
Preterm,[b] %	6.45	10.17	11.09	<.01
Very low birthweight,[c] %	0.66	0.86	1.58	<.01
Low birthweight,[d] %	4.66	6.44	8.39	<.01
Heavy birthweight,[e] %	5.61	2.01	2.08	<.01
Low birthweight and preterm,[f] %	2.49	3.35	4.96	<.01
Low birthweight and term,[g] %	2.11	2.92	3.27	<.01
Neonatal mortality rate	4.7	5.3	6.3	.36
Postneonatal mortality rate	2.1	2.2	3.5	.19
Infant mortality rate	6.8	7.5	9.7	.09
Mean birthweight (g)	3432	3229	3206	<.01
Mean gestation (wk)	39.86	39.07	39.08	<.01

Note. Data are for single live births to adult married, military-dependent mothers residing in Hawaii 1979 through 1989.
[a]Less than 33 weeks of gestation.
[b]Less than 37 weeks of gestation.
[c]Less than 1500 g.
[d]Less than 2500 g.
[e]Greater than or equal to 4250 g.
[f]Less than 2500 g and <37 weeks of gestation.
[g]Less than 2500 g and ≥37 weeks of gestation.

vantages. Hawaii has a multiethnic culture without a single predominant ethnic group. It has a relatively stable and temperate climate. There is a sizable US military population that has access to a large military health care center (Tripler Army Medical Center), where perinatal health care with presumably consistent content and quality is available to eligible recipients. Although these features do not provide precise control for all disparities in living conditions, employment status, racism, environment, and access to and use of health care resources, they offer modest advantages in an effort to reduce the influence of these potential confounders.

Methods

The data for this study were obtained from the 1979–1989 Hawaii vital-record file that provides linked live birth–infant death information. This file contains coded information from all certificates of live birth from the study period and appended information from the death certificates of those infants who died during the first year of life. Although the state of Hawaii has reciprocal vital-record agreements with other US states, no studies were located that investigated whether an appreciable underreporting of out-of-state infant deaths occurs or whether underreporting varies by the military status and ethnicity of the mother. Single live births to mothers residing in Hawaii were selected for analysis based on the following additional criteria: (1) the mother was married, 18 years of age or older, and non-Hispanic White, Black, or Filipino, based on the stated race of the mother on the birth certificate; and (2) the birth certificate indicated that the father was on active-duty status in the US military.

The χ^2 test and *t* test were used to test for significant bivariate ethnic differences in maternal characteristics and pregnancy outcomes. Multiple logistic regression was used to calculate odds ratios for the independent effects of maternal factors on low birthweight and neonatal mortality.

Gestational age in completed weeks was calculated as the interval between the date of last normal menses and the date of birth, with the use of recommended procedures for imputing values when the day of last normal menses is unknown.[24] Adequacy of prenatal-care utilization was determined based on a modification of the Institute of Medicine index.[20]

Results

Maternal and paternal sociodemographic, prenatal-care utilization, and hospital utilization characteristics are shown by ethnic group in Table 1. Significant ethnic differences were observed for every factor examined. Compared with Black and White mothers, Filipino mothers were more likely to be 35 years of age and older, were less likely to initiate prenatal care in the first trimester, and were more likely to have less than 12 years of education. Lower percentages of inadequate prenatal-care utilization and high parity for age were found for White mothers. The lowest proportion of fathers with more than 12 years of education was observed for Blacks.

Table 2 presents pregnancy outcome data by ethnicity. Significant ethnic differences were apparent for all measures of birthweight and gestational age. The mean birthweights and gestational ages of Black and Filipino infants were similar, but these were approximately 200 g and three quarters of a week less than the mean birthweight and gestational age, respectively, of White infants. White infants had the lowest percentages of low birthweight and preterm delivery. Black and Filipino infants had similar percentages of preterm delivery, but Black infants had markedly higher percentages of very low birthweight and very preterm delivery.

No significant ethnic differences in neonatal, postneonatal, and infant mortal-

ity rates were observed. White infants had the lowest rates and Black infants had the highest rates of neonatal, postneonatal, and infant mortality.

Birthweight percent distributions by maternal ethnic group are provided in Figure 1. The Black and Filipino distributions were similarly positioned on the birthweight axis but shifted relative to the White distribution toward the lower end of the birthweight scale. The low-birthweight portion of the distribution is highlighted in the left-hand portion of the figure. Below 1750 g, the Filipino and White distributions are more similar to each other, with the Black distribution having a relatively greater percentage of infants weighing less than 1500 g (see Table 2). Between 1750 and 2500 g, the birthweight percent distributions for Blacks and Filipinos are more similar, with lower proportions observed for Whites. Ethnic variations were observed in the gestational-age distribution and followed patterns similar to those observed for birthweight.

The results of the logistic regression analyses, focusing on very low and low birthweights, are presented in Table 3. The same independent characteristics considered in each analysis included measures of maternal age, parity, maternal and paternal education, prenatal-care utilization, hospital of delivery, and ethnicity. In both analyses, older maternal age (≥35 years), primiparity, inadequate prenatal care, and a nonmilitary hospital of delivery were identified as significant risk factors. A father with more than 12 years of education was found to be a protective factor, reducing the risk of these outcomes. Both Filipino and Black maternal ethnicity increased the risk of low birthweight compared with the risk for Whites. For very low birthweight, the Black odds ratio (2.48, confidence interval [CI] = 1.87, 3.28) was significantly higher than that for Whites, but the Filipino odds ratio (1.25, CI = 0.88, 1.78) was not significantly different than that for the White reference group.

Table 4 presents the results of logistic regression analyses that focused on neonatal and infant mortality. The same independent variables used in the preceding analyses (Table 3) were considered. Primiparity, inadequate prenatal care, and a nonmilitary hospital of delivery increased the risk of both mortality measures, whereas a father with more than 12 years of education decreased the risk. Maternal age of 35 years or more was found to increase the risk of infant mortality. Maternal ethnicity was not found to significantly increase the risk of neonatal mortality. A significantly higher risk of infant mortality was identified only for Black mothers.

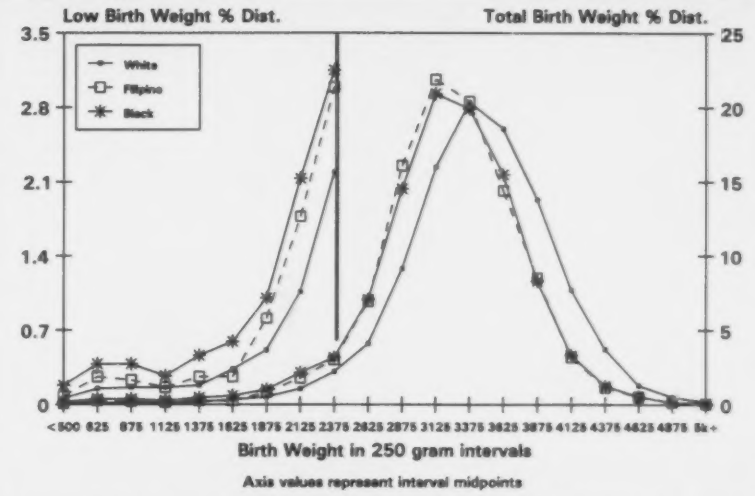


FIGURE 1—Birthweight distribution by maternal ethnicity, Hawaii resident live births to adult married military dependents, 1979 through 1989.

TABLE 3—Results of Logistic Regression Analysis of Low Birthweight and Preterm Delivery

Risk Factor	Low Birthweight			Very Low Birthweight		
	Odds Ratio	95% CI	*P*	Odds Ratio	95% CI	*P*
≥35 years of age	1.50	1.18, 1.90	<.01	2.16	1.26, 3.69	<.01
Primipara	1.41	1.28, 1.56	<.01	1.70	1.33, 2.17	<.01
High parity for age[a]	1.86	1.16, 2.99	.01	2.47	0.89, 6.86	.08
<12 years of education (mother)	1.17	1.00, 1.37	.06	0.88	0.56, 1.37	.56
>12 years of education (father)	0.85	0.77, 0.94	<.01	0.71	0.55, 0.93	.01
Inadequate or no prenatal care[b]	1.78	1.46, 2.16	<.01	2.19	1.42, 3.36	<.01
Nonmilitary hospital delivery	1.28	1.14, 1.44	<.01	1.79	1.36, 2.35	<.01
Filipino mother	1.37	1.20, 1.57	<.01	1.25	0.88, 1.78	.22
Black mother	1.91	1.69, 2.15	<.01	2.48	1.87, 3.28	<.01

Note. Data are for single live births to adult, married, military-dependent mothers residing in Hawaii 1979 through 1989.
[a]High parity for age: for women 18–21 years, three or more births; for women 22–24 years, four or more births; for women 25+ years, five or more births.
[b]Institute of Medicine Index of Adequacy of Prenatal Care.[20]

To explore whether these observed ethnic patterns in the risk of mortality varied by birthweight, the study population was divided into low-birthweight (<2500 g) and normal-birthweight (≥2500 g) groups. Infant mortality was again investigated for each birthweight group by use of logistic regression that contained the same independent variables indicated in Table 4. No significant differences in the risk of infant mortality were found among the three ethnic groups. For low-birthweight infants, the odds ratio for the Black group was 1.07 (CI = 0.70, 1.64), and the odds ratio for the Filipino group was 0.86 (CI = 0.53, 1.41). For normal-birthweight infants, the odds ratio for Blacks was 0.84 (CI = 0.45, 1.56), and the odds ratio for Filipinos was 0.91 (CI = 0.49, 1.70).

The six categories proposed by Baruffi et al. (perinatal-related conditions, injuries, infections, sudden infant death syndrome, congenital anomalies, other)[25] were used to examine causes of infant death for each maternal ethnic group. The small number of cases in each category precluded an examination of significant differences; however, higher proportions of deaths due to sudden infant death syndrome and perinatal-related conditions were observed for Blacks.

Percentages of low and very low birthweights, percentages of preterm and very preterm deliveries, and neonatal and infant mortality rates were examined

TABLE 4—Results of Logistic Regression Analysis of Neonatal and Infant Mortality

Risk Factor	Neonatal Mortality			Infant Mortality		
	Odds Ratio	95% CI	P	Odds Ratio	95% CI	P
≥35 years of age	1.67	0.79, 3.50	.18	1.97	1.06, 3.65	.03
Primipara	1.42	1.04, 1.94	.03	1.33	1.01, 1.74	.04
High parity for age[a]	2.03	0.48, 8.48	.33	2.70	0.96, 7.59	.06
<12 years of education (mother)	1.00	0.59, 1.70	.99	1.16	0.76, 1.77	.50
>12 years of education (father)	0.65	0.46, 0.92	.01	0.66	0.49, 0.89	<.01
Inadequate or no prenatal care[b]	1.92	1.06, 3.43	.03	1.93	1.18, 3.17	<.01
Nonmilitary hospital delivery	1.67	1.17, 2.37	<.01	1.51	1.11, 2.06	<.01
Filipino mother	1.10	0.70, 1.71	.69	1.07	0.73, 1.56	.73
Black mother	1.35	0.88, 2.04	.15	1.52	1.08, 2.16	.02

Note. Data are for single live births to adult, married, military-dependent mothers residing in Hawaii 1979 through 1989.
[a]High parity for age: for women 18–21 years, three or more births; for women 22–24 years, four or more births; for women 25+ years, five or more births.
[b]Institute of Medicine Index of Adequacy of Prenatal Care.[20]

TABLE 5—Pregnancy Outcomes, by Ethnicity of Mother

Outcome	Ethnicity of Mother, %		
	White (n = 28 046)	Filipino (n = 22 309)	Black (n = 716)
Very preterm[a]	1.05	1.88	3.17
Preterm[b]	5.90	10.36	11.40
Very low birthweight[c]	0.62	0.81	2.94
Low birthweight[d]	4.19	8.26	9.37
Neonatal mortality rate	3.6	5.1	11.2
Infant mortality rate	5.1	7.1	16.8

Note. Data are for single live births to adult, married, non–military-dependent mothers residing in Hawaii 1979 through 1989.
[a]Less than 33 weeks of gestation.
[b]Less than 37 weeks of gestation.
[c]Less than 1500 g.
[d]Less than 2500 g.

for single live births to non–military-dependent, adult, married mothers residing in Hawaii (Table 5). These measures of birth outcomes were calculated for a cursory comparison with those of the military-dependent study population (Table 2). For Whites, each of the above six measures was higher for the military group compared with the nonmilitary population. For Blacks, a relatively small nonmilitary resident population in Hawaii, the reverse was observed, with the military group demonstrating lower percentages and rates than the nonmilitary population. For Filipinos, the military group had lower percentages of preterm deliveries, very preterm deliveries, and low birthweights but had higher neonatal and infant mortality rates.

Discussion

This study population was selected to examine multiethnic differences in pregnancy outcomes among adult, married women whose spouses were members of the US military. These women were presumably not living in extreme poverty and had similar availability of and access to comparable prenatal-care and hospital delivery services. Using the year 2000 target objectives for US health status as a reference standard, the measures of pregnancy outcome for this study population were relatively good and in most cases were near or below the year 2000 target objectives for the respective ethnic groups.[26] (No target objectives have been published for Filipinos.) Nevertheless, notable ethnic disparities in some of the indicators remain.

After controlling for available socioeconomic characteristics, the risk of low birthweight for infants of Black mothers was almost twice that for infants of White mothers, approximating the Black to White ratio of low-birthweight percentages found in many other studies.[1,5,8,22] Compared with infants of Black mothers, the infants of Filipino mothers were shown to have a lower percentage of low birthweight, although their mean birthweight was comparable to that of Black infants. Filipino mothers initiated contact with and utilized prenatal-care services slightly less adequately than Black mothers. Within low- and normal-birthweight categories, no significant differences were found among the three ethnic groups in infant mortality, suggesting that the observed ethnic variation in overall mortality rates was mainly a function of birthweight distribution differences. In particular, the main difference between the Filipino and Black groups was the strikingly excessive proportion of infants with very low birthweight born to Black mothers.

Nativity status of the mother has been identified by several studies as being associated with pregnancy outcomes.[27–29] In this study, a large proportion of Filipino mothers were foreign-born, whereas the vast majority of White and Black mothers were US-born. A previous investigation of the effect of nativity status on the outcomes of Filipino births in Hawaii indicated few differences between US- and Philippines-born women residing in Hawaii.[30] Due to the high correlation between nativity status and Filipino ethnicity, this variable was excluded from this analysis.

In spite of the availability of and access to health care facilities, the proportion of mothers in the study population who initiated prenatal care in the first trimester did not reach the year 2000 target objective of 90%. Other studies have noted decreased utilization of prenatal-care services among military personnel and have commented that the reduction of financial barriers alone may not be sufficient to stimulate increased utilization of prenatal-care services.[15]

The finding of a higher risk of low birthweight and mortality among infants delivered at a nonmilitary hospital was unexpected. One explanation for these results is that some high-risk cases may have been delivered at civilian tertiary-care centers. This possibility was investigated. It was found that the majority of the nonmilitary hospital deliveries did occur at another tertiary-care center. Unfortunately, these data are insufficient to document that these cases represented high-risk transfers.

Paternal education in excess of 12 years was found to have a protective effect on low birthweight and infant mortality. Paternal education has less frequently been considered in pregnancy outcomes investigations, and among married

women, it may provide a useful additional indicator of socioeconomic status. In this particular population, paternal education may in part reflect military rank with accompanying variations in levels of pay, housing allowance, and social status.

In this study, no measures of maternal smoking, height, weight, and weight gain during pregnancy were available for consideration. Blacks have been reported to smoke less than Whites, and this observation has also been made in military populations.[14,19,21] Variations in maternal height and weight, which have been found to be related to infant birthweight, are likely among these three ethnic groups.[31] Although no information could be identified to suggest how these measures would vary among ethnic groups in the US military population, they constitute an area for further review.

Although subordinate to the primary focus of this paper, ethnic-specific differences in pregnancy outcome measures were observed between the military and nonmilitary resident populations of adult, married mothers (Tables 2 and 5). These cursory findings, suggesting that the White military group may have worse outcomes and the Black military group may have better outcomes than their nonmilitary counterparts, indicate the need for further study of ethnic-specific differences in pregnancy outcomes between military and nonmilitary populations in terms of socioeconomic and employment factors; parental risk characteristics; and sources, management, and utilization of perinatal health care services.

Overall, the findings of this study of ethnic variation in pregnancy outcomes underscore the complex and poorly understood nature of this phenomenon. Although ethnic differences in low-birthweight and very low birthweight percentages persisted at significant levels, the neonatal mortality rates among ethnic groups approached greater comparability. These observations raise further questions regarding the validity of low birthweight as a multiethnic indicator of mortality risk.[32] This issue continues to emerge and has been highlighted by recent disclosures of information on Asian birthweight distributions and infant mortality rates.[33] However, theories that propose the possibility of normal ethnic variations in mean birthweight and gestational age do not explain the excess percentages of very preterm deliveries and very low birthweight found among infants born to Black mothers in this study. The prevention of very preterm birth in Black populations remains a separate and demanding issue, which must be addressed to reduce the ethnic gap in pregnancy outcome indicators found in the United States. □

References

1. Schoendorf KC, Hogue CJR, Kleinman JC, Rowley D. Mortality among infants of Black as compared with White college educated parents. *New Engl J Med.* 1992;326: 1522–1526.
2. Alexander GR, Tompkins ME, Altekruse JM, Hornung CA. Racial differences in the relation of birth weight and gestational age to neonatal mortality. *Public Health Rep.* 1985;100:539–547.
3. Binkin NJ, Williams RL, Hogue CJR, Chen PM. Reducing black neonatal mortality. Will improvement in birth weight be enough? *JAMA.* 1985;253:372–375.
4. Wise PH, Kotelchuck M, Wilson MS, Mills M. Racial and socioeconomic disparities in childhood mortality in Boston. *N Engl J Med.* 1985;313:360–366.
5. Shiono PH, Klebanoff MA, Graubard BI, Berendes HW, Rhoads GG. Birth weight among women of different ethnic groups. *JAMA.* 1986;255:48–52.
6. Geronimus AT. The effects of race, residence, and prenatal care on the relationship of maternal age to neonatal mortality. *Am J Public Health.* 1986;76:1416–1421.
7. Kleinman JC, Kessel SS. Racial differences in low birth weight. Trends and risk factors. *N Engl J Med.* 1987;317:749–753.
8. Alexander GR, Cornely DA. Racial disparities in pregnancy outcomes: the role of prenatal care utilization and maternal risk status. *Am J Prev Med.* 1987;3:254–261.
9. Gould JB, LeRoy S. Socioeconomic status and low birth weight: a racial comparison. *Pediatrics.* 1988;82:896–904.
10. Kessel SS, Kleinman JC, Koontz AM, et al. Racial differences in pregnancy outcomes. *Clin Perinatol.* 1988;15:745–754.
11. Murray JL, Bernfield M. The differential effect of prenatal care on the incidence of low birth weight among blacks and whites in a prepaid health care plan. *N Engl J Med.* 1988;319:1385–1391.
12. Hogue CJR, Yip R. Preterm delivery: can we lower the Black infant's first hurdle? *JAMA.* 1989;262:548–550.
13. Emanuel I, Hale CB, Berg CJ. Poor birth outcomes of American Black women: an alternative explanation. *J Public Health Policy.* Autumn 1989;10(3):299–308.
14. Fichtner RR, Sullivan KM, Zyrkowski CL, Trowbridge FL. Racial/ethnic differences in smoking, other risk factors, and low birth weight among low-income pregnant women, 1978–1988, in CDC surveillance summaries. *MMWR.* July 1990;39(No. SS-3):13–21.
15. Kugler JP, Connell FA, Henley CE. Lack of difference in neonatal mortality between Blacks and Whites served by the same medical care system. *J Fam Pract.* 1990; 30:281–288.
16. Hulsey TC, Levkoff A, Alexander GR. Birth weights of infants of Black and White mothers without pregnancy complications. *Am J Obstet Gynecol.* 1991;164:1299–1302.
17. Hulsey TC, Levkoff A, Alexander GR, Tompkins ME. Differences in Black and White infant birth weights: the role of maternal demographic factors and medical complications of pregnancy. *South Med J.* 1991;84:443–446.
18. Rawlings JS, Weir MR. Race- and rank-specific infant mortality in a U.S. military population. *Am J Dis Child.* 1992;146:313–316.
19. Adams ME, Read JA, Rawlings JS, Harlass FB, Sarno AP, Rhodes PH. Preterm delivery among Black and White enlisted women in the United States Army. *Obstet Gynecol.* 1993;81:65–71.
20. Kessner DM, Singer J, Kalk CW, et al. Infant death: an analysis of maternal risk and health care. In: *Contrasts in Health Status.* Washington, DC: Institute of Medicine, National Academy of Sciences; 1973;1:1–203.
21. Kogan MD, Kotelchuck M, Alexander GR, Johnson W. Racial disparities in reported prenatal care advice from health care providers. *Am J Public Health.* In press.
22. Kleinman JC. Infant mortality among racial/ethnic minority groups, 1983–1984, in CDC surveillance summaries. *MMWR.* July 1990;39(No. SS-3):31–39.
23. Greenberg DN, Yoder BA, Clark RH, Butzin CA, Null DM. Effect of maternal race on outcome of preterm infants in the military. *Pediatrics.* 1993;91:572–577.
24. Alexander GR, Tompkins ME, Cornely DA. Gestational age reporting and preterm delivery. *Public Health Rep.* 1990;105:267–275.
25. Baruffi G, Alexander GR, Novotny R. Causes of infant mortality in a multi-ethnic population. *Asia Pac J Public Health.* 1990;4(2–3):145–150.
26. *Healthy People 2000. National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50213.
27. Ventura S, Taffel S. Childbearing characteristics of US and foreign-born Hispanic mothers. *Public Health Rep.* 1985;100: 539–547.
28. Cabral H, Fried LE, Leverson S, Amaro H, Zuckerman B. Foreign-born and US-born black women: differences in health behaviors and birth outcomes. *Am J Public Health.* 1990;80:70–72.
29. Kleinman JC, Fingerhut LA, Prager K. Differences in infant mortality by race, nativity status, and other maternal characteristics. *Am J Dis Child.* 1991;145:194–199.
30. Alexander GR, Baruffi G, Mor J, Kieffer EC. Maternal nativity status and pregnancy outcome among U.S.-born Filipinos. *Social Biol.* 1992;39(3–4):278–284.
31. Kleinman JC, Madans JH. The effects of maternal smoking, physical stature and educational attainment on the incidence of low birth weight. *Am J Epidemiol.* 1985; 121:843–855.
32. Wilcox A, Russell I. Why small black infants have a lower mortality rate than small white infants: the case for population-specific standards for birth weight. *J Pediatr.* 1990;116:7–10.
33. Alberman E. Are our babies becoming bigger? *J R Soc Med.* 1991;84:257–260.

Custody of Cocaine-Exposed Newborns: Determinants of Discharge Decisions

Daniel R. Neuspiel, MD, MPH, Terry Martin Zingman, MPH, Virginia H. Templeton, BA, Paula DiStabile, MSN, JD, and Ernest Drucker, PhD

ABSTRACT

Objectives. Maternal cocaine use is a leading grounds for newborn foster placement. This study was initiated to investigate the factors that predict custody status of infants born to substance-abusing women.

Methods. A retrospective cohort design was used to study the correlates of discharge custody decisions for 99 consecutive infants testing positive for cocaine in a public hospital.

Results. The population was 49% Black, 40% Hispanic, and 11% other or unknown. Custody at discharge was to mothers (38%), other family members (25%), or agency foster care (36%). Placement outside the family was greater when mothers had prior child welfare records, in Blacks vs others, with no prenatal care, and when mothers were younger at their first delivery or older at the index birth. Denial of custody to the mother was higher with prior child welfare involvement, in Blacks, and when the mother did not live in her own home. Both models also controlled for parity, child sex, and birthweight.

Conclusions. Earlier involvement with child welfare authorities, race, and other factors predict continued separation of mothers and children at newborn discharge, suggesting the need to reexamine current policies and practices. (*Am J Public Health.* 1993;83:1726–1729)

Introduction

Substance use in pregnancy is a major social and medical problem, affecting the woman, her new child, and other family members.[1] Because of concern about the ability of drug-using women to care adequately for their children, child protection authorities in many jurisdictions have acted to separate these mothers from their infants after birth.[2–4]

In the state of New York, child abuse and neglect laws have been applied in cases in which newborn urine assays have shown evidence of illicit drugs, particularly cocaine or its metabolites. Local child welfare authorities in New York City are delegated the authority to investigate such cases and to decide temporary placement at the newborn's hospital discharge. These decisions are made after input by hospital social work staff and assessment of the family and home environment. Some of these infants are discharged to their mothers or other family members, and some are discharged to foster care agencies for nonfamily placement. On occasion, mothers may voluntarily agree to have their infant placed in foster care. Maternal cocaine use has become the leading grounds for newborn foster placement in New York City. Substance use also contributed to the 29% nationwide increase in foster care between 1986 and 1989.[5]

In New York State, immediate newborn discharge placement is followed by a family court determination of custody and may not represent the long-term location of custody. However, separation of mother and child at this time may have a lasting impact on the mother–infant relationship.[6] Prior studies have not reported on factors influencing the decision to send a cocaine-exposed newborn home with his or her mother or to foster care placement with a relative or agency. Also, no previous research has studied the predictors of foster care placement of infants of substance-abusing mothers. Although child welfare policy calls for reunifying separated mothers and children, little information exists on the success of current policies in promoting family cohesion. This study was performed to explore the determinants of newborn discharge custody placement in a cohort of cocaine-exposed infants.

Methods

Subjects

This investigation used a retrospective cohort design. The study sample included all newborns testing positive for cocaine or its metabolite benzoylecgonine by enzyme-immunoassay-technique urine assay[7] from July 1990 through May 1991 at a public hospital in New York City. During this period, there were 3261 births at the hospital. Cocaine toxicology tests were ordered selectively in newborns with at least one of the following criteria: (1) no documented prenatal care, (2) known maternal history of illicit drug use, (3) maternal or newborn signs or symptoms consistent with drug exposure, or (4) birth outside the hospital. The study period im-

The authors are with the Department of Epidemiology and Social Medicine, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY. Daniel R. Neuspiel is also with the Department of Pediatrics, Albert Einstein College of Medicine and Montefiore Medical Center, and the Department of Ambulatory Care, North Central Bronx Hospital.

Requests for reprints should be sent to Daniel R. Neuspiel, MD, MPH, NCBH 4M-08, 3424 Kossuth Ave, Bronx, NY 10467.

This paper was accepted April 29, 1993.

mediately preceded the implementation of a modified child welfare investigation and placement policy in New York State in June 1991. The new policy eliminated lack of prenatal care and extrahospital births as criteria for newborn toxicology.

Data Collection

Research staff gathered all information by systematic review of maternal and infant medical records using prepared coding forms. The chart reviewers were not aware of the specific goal of this investigation.

Statistical Analysis

Chi-square tests (for categorical variables) and *t* tests (for continuous, normally distributed variables) were used in conducting bivariate comparisons. Ordinal grouped variables were examined with the Kruskal-Wallis one-way analysis of variance. Multiple logistic regression was used to control simultaneously for several independent variables predicting discharge custody status. SAS software was used for all analyses.[8] Methods described by Schlesselman[9] were used to determine confidence intervals on multiple logistic odds ratios.

Results

During the study period, 99 infants were identified with positive urine assays for cocaine or benzoylecgonine. Reasons for toxicology testing (not mutually exclusive) were known history of drug use (64), clinical signs of drug use in mother or newborn (6), lack of prenatal care (31), and extramural delivery (11). Custody at newborn discharge (see Table 1) was to birth mothers (38%), other family members (25%), or agency foster care (36%). No differences were noted in reasons for testing between these custody groups.

Mothers who retained their infant's custody at discharge were younger at the index childbirth, although their first deliveries occurred at an older age (both *P*s < .05). Mothers who kept their newborns were also less often Black ($P < .01$). They were more likely to live in their own home and to have fewer previous children and any prenatal care (all *P*s < .05). Education level (high school completion) did not differ by discharge status or ethnicity. Reported history of substance use did not differ in the three custody groups.

Of mothers with previous children, those who retained custody at discharge were less likely to have had previous reporting to the local or state child welfare agency ($P < .0005$; Table 2). Index newborns with prior siblings in agency or family foster care were more likely to have been discharged to nonmaternal care. Among 50 mothers with prior child welfare records (range = 0 to 9 children, median = 2 children per mother), 109 prior children were known to have been placed out of maternal care in the past, and 107 children were placed out of maternal care at the time of the index birth. The number of prior children placed out of maternal care was higher among mothers whose index children were not discharged to their custody ($P = .095$, Kruskal-Wallis test). Newborn characteristics of the study subjects (Table 3) did not differ by discharge custody status.

TABLE 1—Characteristics of Mothers of 99 Cocaine-Positive Newborns, by Discharge Custody Status

	Discharge Status		
	Mother (n = 38)	Other Relative (n = 25)	Agency (n = 36)
Mean maternal age, y (SD)	26.8 (5.3)	27.2 (4.3)	29.8 (4.6)[a]*
Mean maternal age at first birth, y (SD)	20.4 (3.9)	20.1 (3.2)	18.3 (4.1)**
Ethnicity, no. (%)***			
Hispanic	23 (61)	8 (32)	9 (25)
Black	10 (26)	14 (56)	24 (67)
White	3 (8)	2 (8)	2 (6)
Unknown	2 (5)	1 (4)	1 (3)
Education level, no. (%)			
≥12 y	17 (45)	7 (28)	11 (31)
<12 y	9 (24)	6 (24)	14 (39)
Unknown	12 (32)	12 (48)	11 (31)
Lives in own home, no. (%)[b]*	23 (61)	11 (44)	11 (31)
One or more prenatal visit, no. (%)[b]*	31 (82)	20 (80)	20 (56)
Multiparous, no. (%)[b]*	28 (74)	24 (96)	33 (92)

[a]*t* test: agency vs mother and agency vs relative.
[b]Chi-square test.
*$P < .05$.
**$P = .04$ (*t* test: agency vs mother).
***$P = .01$ (chi-square test excluding unknowns).

TABLE 2—Child Welfare History of Mothers of 99 Cocaine-Positive Newborns, by Discharge Custody Status

	Discharge Status		
	Mother (n = 38)	Other Relative (n = 25)	Agency (n = 36)
Prior child welfare record,[a] no. (%)	5 (13)	13 (72)	27 (75)
No prior record, no. (%)	23 (61)	6 (24)	6 (17)
First child, no. (%)	10 (26)	1 (4)	3 (8)
Prior child placed in foster care,[a] no. (%)	2 (5)	13 (52)	21 (58)
Prior children in foster care,[a] median no. (range)	0 (0–2)	1.5 (0–7)	2 (0–9)

[a]Agency or family foster care.
*$P < .0005$ (chi-square test comparing groups with previous children).

Two models were used in performing a multiple logistic regression analysis (Table 4): (1) agency (nonfamily) placement vs mother or other relative and (2) nonmaternal placement (relative or agency) vs mother. Independent dichotomous (prior child welfare involvement, Black vs non-Black, no prenatal care, not living in own home, primiparous, and newborn gender) and continuous (maternal age at index and first birth and newborn weight) variables were entered simultaneously in both models.

Previous involvement of the mother with the child welfare agency was the strongest predictor of nonmaternal discharge placement; it was less strongly related to nonfamily placement. Black race

TABLE 3—Neonatal Characteristics of 99 Cocaine-Positive Newborns, by Discharge Custody Status

	Discharge Status			
	Mother (n = 38)	Other Relative (n = 25)	Agency (n = 36)	Total (n = 99)
Female gender, no. (%)	15 (39)	12 (48)	22 (61)	49 (49)
Birthweight <2500 g, no. (%)	10 (26)	6 (25)	11 (31)	27 (27)
Mean birthweight, g (SD)	2865 (486)	2748 (693)	2669 (566)	2764 (572)
Mean birth length, cm (SD)	48.1 (2.9)	48.2 (4.0)	47.2 (3.9)	47.8 (3.5)
Mean birth head circumference, cm (SD)	32.4 (2.5)	32.4 (1.8)	32.6 (2.9)	32.4 (2.5)
Mean gestational age, wk (SD)	38.6 (1.7)	37.3 (3.0)	38.0 (2.3)	38.1 (2.3)

TABLE 4—Predictors of Discharge Custody of 99 Cocaine-Positive Newborns: Multiple Logistic Regression

	Model 1[a]: Nonfamily Discharge		Model 2[b]: Nonmaternal Discharge	
Independent Variable	Odds Ratio	95% Confidence Interval	Odds Ratio	95% Confidence Interval
Prior child welfare involvement	1.81	0.95, 3.43	3.74	1.91, 7.34
Black vs non-Black	1.78	0.98, 3.25	2.04	1.08, 3.86
No prenatal care	2.02	1.10, 3.70	1.06	0.54, 2.11
Not living in own home	1.68	0.94, 3.00	2.56	1.25, 5.24
Female child	1.52	0.85, 2.70	1.39	0.73, 2.65
First child	0.50	0.17, 1.43	1.84	0.65, 5.22
Maternal age, y[c]				
At first child	0.76	0.62, 0.93	0.95	0.79, 1.13
At index delivery	1.16	1.03, 1.31	0.97	0.85, 1.11
Newborn weight, g[c]	1.00	...	1.00	...

[a]Odds ratio of risk for agency placement vs mother or family, controlling for other variables.
[b]Odds ratio of risk for agency or family placement vs mother, controlling for other variables.
[c]Continuous independent variable (odds ratio of the change in risk with change of one unit of independent variable in same direction).

was also predictive of discharge away from the mother and less predictive of nonfamily discharge custody. In neither model were any interactions noted between ethnicity and prior child welfare reporting. In addition, lack of prenatal care predicted nonfamily placement, and the mother's lack of her own home was associated with her not gaining custody of her newborn at discharge. Lower age of the mother at first delivery and older age at delivery of the index newborn were also predictive of nonfamily placement.

Discussion

The discharge custody status of newborns known to be cocaine exposed may be influenced by many factors, including local policies and practices,[3] individual differences influencing parenting ability and the home environment, and the availability of a family support network. Some of these placements may have been emergency dispositions prior to final court action and may not necessarily reflect court-ordered custody decisions. Others may have been voluntary placements by mothers hoping for judicial leniency in final custody decisions.

Federal statutes require that, before the placement of a child in foster care, "reasonable efforts" be made to prevent the need for removal and to enable the child to return to his or her home.[10] In New York, state social service statutes call for families to be kept together whenever possible, not to be separated solely because of poverty, and to be offered services to "maintain and strengthen family life."[11] These statutes affirm "the state's first obligation" as assisting "the family with services to prevent its break-up or to reunite it" when the child is separated.[12,13] Unnecessary protracted foster care "may deprive these children of positive, nurturing family relationships and have deleterious effects on their development."[14] Courts may order agencies to "encourage and strengthen" the parental relationship, including helping parents with "housing, employment, counselling, medical care or psychiatric treatment."[15,16] Also, courts and parents may agree to a temporary separation from the child with plans to reunite after successful drug treatment.

The role of prior child welfare involvement in increasing the likelihood of removal of subsequent newborns from the home raises questions about compliance with federal and New York State regulations. These statutes explicitly require reasonable efforts for family unification, while early newborn placement may result in long-term separation of children from their birth families. Although repeated removal of newborns from the same addicted mother may reflect her lack of response to drug treatment or other services, it may also indicate the system's failure to individualize assessments and to adequately provide needed services. An overtaxed child welfare system with limited resources may focus on short-term goals of perceived child protection rather than the broader efforts needed to protect children by strengthening and preserving family units.

Reunification of foster children with their birth parents has been supported because (1) the child's well-being is enhanced by continuous caretaking, (2) the birth parents have legal and social priority as permanent caregivers, (3) the child may be psychologically harmed as a result of the separation of foster care, and (4) the state avoids the cost and recruitment of foster parents.[17] But in an era of shrinking resources for family support, limited drug treatment services for women with children,[18] and staff cutbacks, such reunification is more difficult to achieve. In addition, child welfare agencies are under legal pressure to make rapid determinations of newborn custody status.

> Families that could be helped are all too often simply written off, as yet another child begins a career of foster placement. In the name of protection, crimes are being committed not only against children but against parents, who are

left without authority or responsibility and whose only remaining power is to create more children.[19]

In this context, previous records of reporting to these authorities, rather than triggering assistance in family reconstruction, may become a convenient way to identify mothers "at risk" for subsequent parenting deficiency. Overworked and understaffed child welfare agencies may view prior reports as indicators to separate mother and infant until a court determination is made; however, this approach may not be most beneficial to family preservation. The knowledge of cocaine exposure may itself be influenced by social or racial bias in decision making for drug toxicology testing and reporting.[20] Such biases may also have an impact on custody disposition decisions. Furthermore, in a selectively screened population such as ours, newborns with positive cocaine toxicology may differ from the general population of cocaine-exposed newborns at this or other hospitals.

Reasons for ethnic disparity in patterns of newborn custody disposition other than direct bias may pertain. Multigenerational patterns of substance use may differ across ethnic groups. Blacks in the population under study may have had less available extended family support to allow mothers or other family members to retain newborn custody because of the severity of drug addiction, unemployment, homelessness, and premature morbidity and mortality. Access to health and legal services and voluntary foster placement may also vary with ethnicity. Of additional concern in these data are the independent effects of lack of housing and prenatal care on newborn discharge dispositions, since these are important indices of social deprivation. Thus, lack of social resources for housing and accessible health care may predispose to loss of child custody.

The cocaine epidemic of the late 1980s led to an overwhelmed child welfare and foster care system and to a plethora of "boarder babies": newborns housed for weeks and months in hospitals for lack of court-determined discharge placement. As this crisis leveled off, New York state and city social service authorities implemented modified policies concerning neonatal urine toxicology screening (memoranda of June 1991 from the New York State Department of Social Services, the New York City Health and Hospitals Corporation, and the New York City Human Resources Administration–Child Welfare Administration). These changes eliminated lack of prenatal care and extrahospital births as sufficient criteria for newborn toxicology testing. Parental drug use is no longer sufficient to determine child maltreatment, although evidence of such use still triggers a child protective investigation. Time limits were set for family and home investigations after reports to authorities of newborns with positive drug toxicologies. The results of these policy changes are not yet clear.

The medical chart data in this study did not afford complete information to understand fully the determinants of discharge custody decisions. Information was usually lacking on past maternal drug use and treatment, detailed family or home assessments, whether foster placement was voluntary, and specific location of prior foster placements, although child welfare investigators presumably had access to these data. Further research is needed to understand the process of discharge custody decision making in regard to drug-exposed newborns. Prospective evaluation of long-term outcomes of these decisions for children and families needs special attention. In particular, the effects of early separation on mother–infant attachment and maternal self-esteem need elucidation. Most important, studies are needed of interventions to enhance the cohesion and reconstruction of families with substance abuse problems. The societal costs of foster care, family disruption, and untreated chemical dependency are considerable. Removal of the drug-exposed infant from his or her family may not be uniformly beneficial to the child, the family, or society. □

Acknowledgments

This research was supported by grant 1-R18-DA06925-01 from the National Institute on Drug Abuse.

This research was presented at the meeting of the Society for Pediatric Research, May 1992, Baltimore, Md.

We thank Penny Demas, June Greene, and Hadassa Padilla for assistance in data collection.

References

1. Neuspiel DR, Hamel SC. Cocaine and infant behavior. *J Dev Behav Pediatr.* 1991; 12:55–64.
2. Connolly WB, Marshall AB. Drug addiction, pregnancy, and childbirth: legal issues for the medical and social services communities. *Clin Perinatol.* 1991;18:147–186.
3. Caldwell MB, Mascola L, Smith W, et al. Biologic, foster, and adoptive parents: care givers of children exposed perinatally to human immunodeficiency virus in the United States. *Pediatrics.* 1992;90:603–607.
4. Larson CS. Overview of state legislative and judicial responses. *Future Child.* 1991; 1:72–84.
5. *Drug-Exposed Infants: A Generation at Risk.* Washington, DC: US General Accounting Office; 1990.
6. Vietze P, Falsey S, Sandler H, et al. Transactional approach to prediction of child maltreatment. *Infant Ment Health J.* 1980; 1:248–261.
7. Jatlow P. Cocaine: analysis, pharmacokinetics, and metabolic disposition. *Yale J Biol Med.* 1988;61:105–113.
8. *SAS/STAT User's Guide, Release 6.03 Edition.* Cary, NC: SAS Institute; 1988.
9. Schlesselman JJ. *Case-Control Studies.* New York, NY: Oxford University Press; 1982:247–248.
10. 42 USCA §671(a)(15) (West 1991).
11. NY Soc Serv Law §131(3) (McKinney 1992).
12. NY Soc Serv Law §384-b(1)(a)(iii) (McKinney 1992).
13. NY Fam Ct Act §1022(a) (McKinney Supp 1993).
14. NY Soc Serv Law §384-b(1)(b) (McKinney 1992).
15. NY Soc Serv Law §392(9) (McKinney 1992).
16. NY Fam Ct Act §1015-a (McKinney Supp 1993).
17. Goerge RM. The reunification process in substitute care. *Soc Serv Rev.* 1990;64:422–457.
18. Chavkin W. Drug addiction and pregnancy: policy crossroads. *Am J Public Health.* 1990;80:483–487.
19. Elizur J, Minuchin S. *Institutionalizing Madness: Families, Therapy, and Society.* New York, NY: Basic Books; 1989.
20. Chasnoff IJ, Landress HJ, Barrett ME. The prevalence of illicit drug or alcohol use during pregnancy and discrepancies in mandatory reporting in Pinellas County, Florida. *N Engl J Med.* 1990;322:1202–1206.

Home Health Clients: Characteristics, Outcomes of Care, and Nursing Interventions

ABSTRACT

Objectives. The purpose of the study was to provide descriptive data about the characteristics of home health clients, the services that nurses provide, and the outcomes of those services. Such data have been sparse.

Methods. This study examined 2403 home health clients served by four agencies in Nebraska, New Jersey, and Wisconsin. Demographic, health history, and clinical data were analyzed. The Omaha System was used as the model for describing and measuring data specific to clients' health-related problems, nursing interventions, and outcomes of care.

Results. The median age of home health clients was 68.6 years. Nurses conducted 70% of all home visits, identified 9107 client problems, and provided over 96 000 interventions. Between admission and dismissal, clients improved by at least 0.52 point on three 5-point problem-specific outcome subscales (Knowledge, Behavior, and Status).

Conclusions. These data show important characteristics of home health clients in a large national sample. They also support the usefulness of the Omaha System in describing and quantifying nursing practice in the community health setting. The magnitude of positive client change between admission and dismissal suggests that community health services do make a difference. (*Am J Public Health.* 1993;83:1730–1734)

Karen S. Martin, MSN, RN, Nancy J. Scheet, MSN, RN, and Mary Ruth Stegman, PhD, RN

Introduction

What are the characteristics of home health clients and the services they receive in today's complex milieu? What effect do nursing services have on those clients? In this article we report the findings of a recent Visiting Nurse Association of Omaha research project that used three classification schemata for examining client–nurse interactions. The schemata are referred to as the Omaha System.

A variety of client characteristics and trends are evident in the health care literature. One of the most extensive documents about current and anticipated health-related problems of the US population is *Healthy People 2000.*[1] That publication, however, emphasizes medical diagnoses; it addresses home health issues, the effect of nurses, and nursing-focused client concerns in a relatively limited manner. Research conducted in community health settings contributes to the understanding of home health clients. Three nationally reported home-care client studies documented the increasing severity of illness and decreasing reimbursement sources. However, the studies emphasized medical diagnoses rather than nursing services.[2–4]

Nurses, the primary providers of home health services, have designed few studies involving home health clients and nursing practice. Rarely have they systematically explored or quantified physical, environmental, and social characteristics of home health clients and the services they receive.[5–7] Few nurse researchers have investigated the effects of health care services on clients, especially on home health clients.[8–11] Increasingly, however, all health care providers are receiving pressure from reimbursement sources, accreditors, and legislators to consider the outcome effectiveness of client services.[12–15] Establishing the relationship between nursing services and client progress is an important element of a minimum data set useful to the nursing profession.[16] The purpose of a nursing minimum data set is to provide an abstraction tool for standardized, comparable, and essential data. Williams[17] noted the importance of a nursing minimum data set but cautioned that pilot tests were needed, especially tests involving public health and home health settings.

Noteworthy for nurses is a compilation of studies edited by Waltz and Strickland.[18–21] It includes three studies that involve client outcome measurement instruments for community health nurses. Nursing diagnoses were used as the framework for home health measurement tools in two of the studies. Nursing diagnoses were based on the Omaha Problem Classification Scheme[22] and on North American Nursing Diagnosis Association vocabulary.[23]

Methods

The Omaha System research project, carried out from 1989 through 1983, was designed to accomplish multiple purposes. In this article we report the frequency and type of the clients' problems; the clients' problem-specific knowledge, behavior, and status outcome changes; and the frequency and type of nursing in-

At the time of the study, Karen S. Martin and Nancy J. Scheet were with the Visiting Nurse Association of Omaha, Omaha, Neb. Mary Ruth Stegman is with Creighton University, Omaha.

Requests for reprints should be sent to Karen S. Martin, MSN, RN, 2115 S 130th St, Omaha, NE 68144.

This paper was accepted April 9, 1993.

terventions occurring in diverse home health agencies. Elsewhere we describe research findings in addition to the client demographics and service characteristics noted in this article; the conceptual framework and structure of the Omaha System; previous research; and the system's applicability, reliability, validity, and efficiency.[24-26]

The Omaha System has been described as a comprehensive schema of nursing diagnoses, client outcomes, and nursing interventions.[25,27,28] The three schemata were developed during a series of research studies conducted between 1975 and 1986. The Omaha System provides a framework for merging a practice and documentation system and offers a method of organizing client and provider data and integrating them into a manual or automated management information system.

Omaha System Schemata

The first component of the system is referred to as the Problem Classification Scheme; it is a client-focused taxonomy of nursing diagnoses. The most global level of the scheme consists of domains called Environmental, Psychosocial, Physiological, and Health Related Behaviors. At more specific levels, these domains comprise 40 client problems, each with a cluster of signs and symptoms. One of the problems from the Physiological domain is circulation; it was the most frequently occurring problem in this study.

The second component of the Omaha System, the Problem Rating Scale for Outcomes, consists of three 5-point Likert-type subscales. The Knowledge, Behavior, and Status subscales offer a numeric measurement framework with which to evaluate changes in the clients' knowledge, behavior, and status in relation to specified health-related problems and selected time frames. At admission to service, a client who exhibits a circulation impairment problem may have a status rating of 2 as evidenced by pitting edema of the ankles and lower extremities. Ideally, with nursing care, client involvement, and a medical plan of care, at dismissal the client may have no edema at dismissal and therefore a status rating of 5.

The third component of the Omaha System is an Intervention Scheme. This is a hierarchical organization of nursing actions that is designed to describe and quantify problem-specific plans and interventions. At the most global level, the Intervention Scheme comprises four broad categories of interventions: Health Teaching, Guidance, and Counseling; Treatments and Procedures; Case Management; and Surveillance. More specific targets or objects of nursing actions and client-specific information are used in conjunction with the categories. For the client who has a circulation impairment problem, the nurse is likely to (1) make a home visit, complete a physical assessment, and provide care and (2) document the intervention category referred to as Health Teaching, Guidance, and Counseling; the target of cardiac care; and specific details about fluid restrictions and potentially contributing factors such as smoking and stress.

Sample

Four home care agencies were purposively selected for diversity in relation to their history of using the Omaha System, size, organization, length of operation, and geographic location. Client records were included in the sample if the clients were admitted and dismissed during each agency's 18-month data collection period, had at least one physiological problem, and received at least two home visits from a nurse.

Home health nurses who visited clients and documented services in client records participated in the study if they had been employed at the agency for at least 3 months, used the Omaha System to document client services, and participated in the initial 4-hour orientation session. All nurses who attended were introduced to the Omaha System schemata and observed a videotape simulating a nurse–client home visit. Working independently, they identified and recorded the client's problems, outcome scale ratings, and interventions.

Data Collection

A two-part client data collection form was developed, pilot tested, and revised. The first part addressed demographic and health history information, including payment source. Service data included the first and last visit dates and number of home visits made by various types of caregivers. The second part included specific service data based on the three schemata of the Omaha System. Data collection focused on client problems, problem-specific knowledge, behavior, and status outcome changes, and interventions provided by various community health staff members.

Procedures

Project personnel who met specific criteria were selected to serve as data abstractors. Criteria included Omaha System expertise, community health experience, and demonstrated proficiency in documentation. Preliminary reliability of project personnel was established prior to data collection and visits to the four sites were scheduled every 3 to 4 months. The data collectors monitored record data to ensure that all staff at the sites were implementing the Omaha System appropriately. They selected client records on the basis of established criteria and abstracted data from these records. Random samples of abstracted data were assessed for reliability throughout the data collection period.

Analysis

Descriptive and inferential data analyses were performed with the Statistical Analysis System (SAS).[29] To maintain confidentiality, each client record was assigned a number and all data were analyzed in the aggregate by agency or total sample. Frequencies and *t* tests were used for the descriptive analysis presented in this article.

Results

Data were collected between October 1989 and October 1991. A total of 2403 clients and their records were admitted into the sample. The client records were obtained from a large urban Nebraska agency (n = 1063), a large urban New Jersey agency (n = 896), a small rural Nebraska agency (n = 263), and a small rural Wisconsin agency (n = 181).

Demographic and Health History

The home health clients served by the four agencies presented a distinctive composite demographic profile. Older, female, married or widowed Caucasian clients dominated the sample. The clients' ages ranged from birth to 104 years; the median age was 68.6 years. Females accounted for 58% (n = 1389) of the sample. Approximately 39% (n = 925) of the sample were married and 35% (n = 850) were widowed. Eighty-five percent (n = 2045) of the clients were Caucasian, 12% (n = 293) were African-American, 2% (n = 43) were Hispanic, and 0.3% (n = 8) were Asian, and another 0.3% (n = 7) were Native American.

The clients' living accommodations, caregiving arrangements, dependency characteristics, and dismissal status were noted. Approximately 40% of the clients lived alone (n = 883), were independent in relation to functional status (n = 973),

TABLE 1—Frequencies of the Most Common Primary Medical Diagnoses of Home Health Clients, According to ICD-9-CM Diagnostic Groups

Diagnostic Group	Frequency (n = 2403)	%
Circulatory system disorders	572	23.8
Neoplasms	406	16.9
Injuries/poisonings	260	10.8
Respiratory system disorders	227	9.5
Endocrine, metabolic, and nutritional disorders	197	8.2
Digestive system disorders	151	6.3
Musculoskeletal system and connective tissue disorders	137	5.7
Ill-defined signs/symptoms	106	4.4
Skin and subcutaneous tissue disorders	82	3.4
Genitourinary system disorders	58	2.4
Other	207	8.6

Note. ICD-9-CM = *International Classification of Diseases*, 9th Revision: Clinical Modification.[30]

TABLE 2—Visits to Home Health Clients, by Caregiving Category

	No. Clients	Visits per Client			Total Visits
		Minimum	Maximum	Mean	
Nursing	2403	2	116	9.11	21 880
Home health aide	634	1	186	9.66	6122
Physical therapy	289	1	61	6.90	1993
Medical social work	249	1	12	1.86	462
Occupational therapy	32	1	29	6.84	219
Speech pathology	19	1	67	12.47	237
Total	2403	2	282	12.87	30 913

Note. Sample selection criteria required a minimum of two nursing visits. Clients could receive visits from one or more types of caregivers.

TABLE 3—Frequency of Problems Experienced by Home Health Clients (n = 2403), by Domain and Agency

Agency	Physiological		Health Related Behaviors		Psychosocial		Environmental		
	Frequency	%	Frequency	%	Frequency	%	Frequency	%	Total
Urban Nebraska	2183	68.3	782	24.5	208	6.5	21	0.7	3194
Urban New Jersey	3014	74.1	909	22.3	88	2.2	56	1.4	4067
Rural Nebraska	567	72.1	139	17.7	70	8.9	10	1.3	786
Rural Wisconsin	722	68.1	271	25.5	60	5.7	7	0.7	1060
All	6486	71.2	2101	23.1	426	4.7	94	1.0	9107

and were dismissed to self-care (n = 930). However, it should be noted that over 50% of the clients were partially dependent (n = 1207) and the remaining 10% were dependent (n = 223). Client dependency was also evidenced by the fact that 59% (n = 1425) of the clients required a spouse or adult child to serve as the primary caregiver and 30% (n = 731) of the clients were dismissed to family or friends.

Reimbursement of services is a critical issue for all home health agencies. Funding sources directly influence the length and amount of service provided to clients. In this study, Medicare was the primary source of reimbursement for 73% (n = 1755) of the clients. The length of service ranged from 1 to 375 days; the mean was 34.74 days.

The primary medical diagnoses of clients were categorized according to ICD-9-CM diagnostic groups.[30] The diagnostic groups that occurred more than 50 times are shown in Table 1; they account for 91.3% of the sample's diagnostic groups. The frequent occurrence of circulatory system disorders and neoplasms, accounting for 40% of the diagnoses, is consistent with national mortality trends. Within the ICD-9-CM diagnostic groups, the three specific medical diagnoses that occurred most frequently were diabetes mellitus, congestive heart failure, and cerebrovascular disease.

Clients' drug profiles were directly related to medical diagnoses. No drugs were prescribed for 67 clients. The 2336 clients who received at least one drug averaged 5.2 drugs per length of service for a total of 12 460 drugs. Cardiac drugs, analgesics and antipyretics, gastrointestinal drugs, diuretics, and central nervous system agents were the most frequently prescribed drugs.

The number of visits made to clients by each of six types of caregivers varied markedly during the clients' length of service (see Table 2). The length of service (range = 1–375 days; mean = 34.74 days) was mentioned above in relation to reimbursement. Nurses conducted 70% of all home visits. Because nurses were the focus of this study, it should be noted that sample selection criteria required a minimum of two nursing visits. Therefore, clients from each agency who received only one nursing visit or only visits from other types of caregivers were not included in the study; a relatively small number of clients were excluded from the research for these reasons. It should also be noted that the sampling plan used in this research reflects the purpose of the study and the number of nurses in the community health practice setting; the plan was not intended to diminish the importance of interdisciplinary contributions to client care.

Omaha System

Based on the sample of 2403 clients and the Problem Classification Scheme, home health nurses identified a total of 9107 client problems or nursing diagnoses. The mean number of problems per client was 3.79; the median was 3 and the mode was 2. Identified problems ranged from 1 to 15 per client. The frequency of problems from the Physiological, Health Related Behaviors, Psychosocial, and Environmental domains for the entire sample and for each agency is depicted in Table 3. For the total sample, problems in the Physiological domain accounted for 71% of all problems. Problems in the Health Related Behaviors domain accounted for 23% of the problems, whereas Psychosocial and Environmental problems accounted for approximately 5% and 1%, respectively. Although inclusion in the sample required that each client have at least one problem in the Physiological domain, the preponderance of physiological problems is typical of home health clients and their reasons for referral. The frequency of occurrence for the total sample and that for each agency were highly similar.

The second component of the Omaha System is the Problem Rating Scale for Outcomes. Table 4 shows the clients'

Knowledge, Behavior, and Status ratings at admission and dismissal for the entire sample and for each agency. It also depicts the average amount of change that occurred during the period of home health nursing service. Clients' Knowledge, Behavior, and Status ratings increased by at least 0.52 point overall. In addition, Table 4 shows the paired *t*-test scores and probability values for the test. The increase in each rating was greater than zero. The magnitude of the *t*-test values demonstrates the overwhelming consistency of clients' improvement in problem-specific knowledge, behavior, and status as they received service from home health nurses and other professionals. It should be noted that every client did not improve, especially on all three subscales. For example, some clients were admitted to an agency's hospice program and were expected to die relatively soon. Their status ratings associated with the primary medical diagnoses were expected to deteriorate.

The third component of the Omaha System is the Intervention Scheme used by nurses to document their actions as they provide nursing service to clients. The order of frequency of occurrence for the four intervention categories is the same for all four agencies: (1) Surveillance; (2) Health Teaching, Guidance, and Counseling; (3) Treatments and Procedures; and (4) Case Management. Table 5 depicts the frequency of interventions for the nine problems that occurred most frequently in the sample: circulation, integument, neuromusculoskeletal function, respiration, medication regimen, nutrition, bowel function, technical procedures, and pain. The order of frequency of occurrence for the four intervention categories was examined. Surveillance was the most frequently occurring intervention category for seven of the nine problems. Health Teaching, Guidance, and Counseling was the primary intervention category for medication regimen problems, and Treatments and Procedures was the most frequent intervention category for technical procedures problems.

TABLE 4—Outcome Ratings on Admission and Dismissal and Mean Increase, by Agency

	Agency				
	Urban Nebraska	Urban New Jersey	Rural Nebraska	Rural Wisconsin	Total
Knowledge rating					
Admission mean (SD)	2.60 (0.65)	2.52 (0.68)	2.65 (0.76)	2.87 (0.54)	2.60 (0.67)
Dismissal mean (SD)	3.18 (0.75)	3.19 (0.82)	3.40 (0.69)	3.54 (0.69)	3.25 (0.79)
Mean increase (SD)	0.58 (0.01)	0.67 (0.01)	0.75 (0.03)	0.67 (0.02)	0.65 (0.006)
t test[a]	45.24	51.71	27.08	32.96	79.91
Behavior rating					
Admission mean (SD)	2.60 (0.66)	2.86 (0.71)	2.75 (0.86)	3.09 (0.65)	2.80 (0.72)
Dismissal mean (SD)	3.16 (0.81)	3.35 (0.84)	3.48 (0.88)	3.60 (0.79)	3.32 (0.84)
Mean increase (SD)	0.56 (0.01)	0.47 (0.13)	0.73 (0.03)	0.51 (0.02)	0.52 (0.006)
t test	43.71	35.78	23.73	23.25	83.97
Status rating					
Admission mean (SD)	2.50 (0.76)	2.84 (0.72)	2.46 (0.84)	3.28 (0.93)	2.74 (0.81)
Dismissal mean (SD)	3.06 (0.95)	3.34 (0.93)	3.38 (0.99)	3.74 (0.97)	3.29 (0.97)
Mean increase (SD)	0.56 (0.01)	0.50 (0.14)	0.92 (0.04)	0.46 (0.02)	0.55 (0.009)
t test	37.43	35.09	25.73	19.13	59.33

Note. In the sample of 2403 clients, 9107 problems were identified.
[a]Paired *t* test of difference between admission and dismissal ratings. For all paired *t* tests, $P \leq .0001$.

TABLE 5—Frequency of Interventions for Selected Problems

	Surveillance		Health Teaching, Guidance, and Counseling		Treatments and Procedures		Case Management		Total
Problem	Frequency	%	Frequency	%	Frequency	%	Frequency	%	Frequency
Circulation	12 459	69.5	4 193	23.4	185	1.0	1 085	6.1	17 922
Integument	7 645	51.0	2 637	17.6	4 120	27.5	581	3.9	14 983
Neuromusculoskeletal function	8 285	67.8	3 135	25.7	123	1.0	668	5.5	12 211
Respiration	7 955	72.1	2 430	22.0	291	2.6	362	3.3	11 038
Medication regimen	3 576	41.4	4 032	46.7	601	7.0	422	4.9	8 631
Nutrition	2 812	49.1	2 466	43.1	360	6.3	84	1.5	5 722
Bowel function	3 358	61.1	1 680	30.6	311	5.7	144	2.6	5 493
Technical procedure	1 309	24.7	1 004	19.0	2 342	44.3	632	12.0	5 287
Pain	3 106	70.1	933	21.1	74	1.7	317	7.1	4 430
All	57 218	59.3	25 335	26.3	8 463	8.8	5 428	5.6	96 444

Note. In the sample of 2403 clients, 9107 problems were identified.

Discussion

This study is unique because it was designed from a nursing perspective and included analysis of an extensive national home health client sample. The analysis demonstrates the complexity of the clients being served by four community health nursing agencies in Nebraska, Wisconsin, and New Jersey. Indeed, the age, caregiving requirements, dependency, and dismissal status of the clients from each agency were very similar. The large number of clients whose services were reimbursed by Medicare further suggests that most clients are homebound, since Medicare regulations require that recipients of service be homebound. The extent of chronic illness experienced by clients in the sample is supported by the most frequently occurring medical diagnoses and the number and type of client problems in the Physiological domain. Client characteristics noted in this study and the dramatic national aging trend confirm the presence of a significant population of frail elderly persons in our society.

The results of this observational study support the usefulness of the Omaha System in describing and quantifying nursing practice. The system provides a reliable and valid method for gathering large quantities of pertinent client problem, nursing intervention, and client outcome data from differing home health agencies.[24–26] Armed with data from individual clients as well as aggregate data, nurses and administrators can use the Omaha System to guide practice, struc-

ture documentation, and organize a client management information system. Such a method reflects the criteria required for a nursing minimum data set.[25]

The ability of the Omaha System to characterize large aggregate populations across different agencies has great potential for future research. Administrators, practitioners, and policymakers can use aggregate data as a basis for examining the client case mix and relating that mix to specific nursing diagnoses, interventions, and client outcomes. Furthermore, the cost-effectiveness of such relationships can be evaluated, an action that is of increasing interest to policy makers. Changes in clients' knowledge, behavior, and status outcome ratings between admission and dismissal in such a large sample strongly suggests that home health service does make a difference. In fact, this is the largest documented observational study that has examined clients' knowledge, behavior, and status both before and after nursing intervention.

Home health nurses can and do provide multidimensional and comprehensive services that they can describe and quantify. They identify and document client problems specific to the referral reason as well as other problems that they consider priorities. Even when Treatments and Procedures was the prescribed intervention, nurses frequently documented Surveillance; Health Teaching, Guidance, and Counseling; and Case Management interventions as well. Such professional practice suggests that nurses recognized the interrelated, complex nature of client problems, the responsibility to judge whether prescribed services were compatible with the home environment, and the need to provide diverse interventions before problem resolution occurred. Nurses should continue to conduct research that addresses client characteristics and the impact of nursing services. Such research is especially timely because of the increasing demand to measure the outcome effectiveness of health-related services. □

Acknowledgments

This work was funded by National Center for Nursing Research (National Institutes of Health) grant 1 R01 NR02192.

This paper was presented at the Seventh World Congress on Medical Informatics, Geneva, Switzerland, September 1992.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services, Public Health Service; 1991. DHHS publication PHS 91-50212.
2. Handy C. Home care of patients with technically complex nursing needs. *Nurs Clin North Am.* 1988;23:315–328.
3. Phillips E, Fisher M, MacMillan-Scattergood D, Baglioni AJ, Torner J. Home health care: who's where? *Am J Public Health.* 1987;77:733–734.
4. Williams B, Phillips E, Torner J, Irvine A. Predicting utilization of home health resources: important data from routinely collected information. *Med Care.* 1990;28: 379–391.
5. Blank J, Atwood J, Longman A. Perceived home care needs of cancer patients and their caregivers. *Cancer Nurs.* 1989;12:78–84.
6. Lancaster J, Lancaster W. Current status of the health care system. In: Stanhope M, Lancaster J, eds. *Community Health Nursing: Processes and Practices.* St. Louis, Mo: CV Mosby Co; 1992:21–33.
7. O'Neill C, Sorensen E. Home care of the elderly: a family perspective. *Adv Nurs Sci.* 1991;13:28–37.
8. Flynn B. Statistics and quantitative methods in community health nursing research. In: Abraham I, Nadzam D, Fitzpatrick J, eds. *Statistics and Quantitative Methods in Nursing.* Philadelphia, Pa: WB Saunders Co; 1989:122–131.
9. Martin K. Research in home care. *Nurs Clin North Am.* 1988;23:373–385.
10. Meisenheimer C. *Quality Assurance for Home Health Care.* Rockville, Md: Aspen Publishers Inc; 1989.
11. Rinke L, Wilson A, eds. *Outcome Measures in Home Care, Vol. 2: Service.* New York, NY: National League for Nursing; 1987.
12. *Report to Congress: Progress of Research on Outcomes of Health Care Services and Procedures.* Rockville, Md: Agency for Health Care Policy and Research; May 1991. AHCPR publication 91-0004.
13. Greenfield S. The state of outcome research: are we on target? *N Engl J Med.* 1989;320:1142–1143.
14. McCloskey J. Implications of costing out nursing services for reimbursement. *Nurs Manage.* 1989;20:44–49.
15. Moritz P. Innovative nursing practice models and patient outcomes. *Nurs Outlook.* 1991;39:111–114.
16. Werley H, Devine E, Zorn C, Ryan P, Westra B. The nursing minimum data set: abstraction tool for standardized, comparable, essential data. *Am J Public Health.* 1991;81:421–426.
17. Williams C. The nursing minimum data set: a major priority for public health nursing but not a panacea. *Am J Public Health.* 1991;81:413–414.
18. Strickland O, Waltz C, eds. *Measurement of Nursing Outcomes, Vol. 2: Measuring Nursing Performance: Practice, Education, and Research.* New York, NY: Springer Publishing Co; 1988.
19. Strickland O, Waltz C, eds. *Measurement of Nursing Outcomes, Vol. 4: Measuring Client Self-Care and Coping Skills.* New York, NY: Springer Publishing Co; 1990.
20. Waltz C, Strickland O, eds. *Measurement of Nursing Outcomes, Vol. 1: Measuring Client Outcomes.* New York, NY: Springer Publishing Co; 1988.
21. Waltz C, Strickland O, eds. *Measurement of Nursing Outcomes, Vol. 3: Measuring Clinical Skills and Professional Development in Education and Practice.* New York, NY: Springer Publishing Co; 1990.
22. Feldman J, Richard R. The measurement of nursing outcomes for home health care. In: Waltz C, Strickland O, eds. *Measurement of Nursing Outcomes, Vol. 1: Measuring Client Outcomes.* New York, NY: Springer Publishing Co; 1988:475–495.
23. Keating S. The measurement of client outcomes in home health care agencies. In: Waltz C, Strickland O, eds. *Measurement of Nursing Outcomes, Vol. 1: Measuring Client Outcomes.* New York, NY: Springer Publishing Co; 1988:457–474.
24. Martin K, Leak G, Aden C. The Omaha System: a research-based model for decision making. *J Nurs Adm.* 1992;22:47–52.
25. Martin K, Scheet N. *The Omaha System: Applications for Community Health Nursing.* Philadelphia, Pa: WB Saunders Co; 1992.
26. Martin K, Scheet N. *The Omaha System: A Pocket Guide for Community Health Nursing.* Philadelphia, Pa: WB Saunders Co; 1992.
27. Barkauskas V. Home health care: responding to need, growth, and cost containment. In: Chaska N, ed. *The Nursing Profession: Turning Points.* St. Louis, Mo: CV Mosby Co; 1990:394–404.
28. Marek K. Classification of outcome measures in nursing care. In: *Classification Systems for Describing Nursing Practice.* Kansas City, Mo: American Nurses' Association; 1988:37–42.
29. *SAS/STAT User's Guide.* Cary, NC: SAS Institute Inc; 1988.
30. *International Classification of Diseases,* 9th Revision: Clinical Modification. Hyattsville, Md: National Center for Health Statistics; 1980. DHHS publication PHS 80-1260.

Cause of Death in Patients with End-Stage Renal Disease: Death Certificates vs Registry Reports

Thomas V. Perneger, MD, MPH, Michael J. Klag, MD, MPH, and Paul K. Whelton, MD, MSc

ABSTRACT

Objectives. The purpose of this study was to assess agreement on cause of death reporting in end-stage renal disease patients by comparing death certificates and reports to an end-stage renal disease registry.

Methods. Death certificates and registry reporting forms were retrieved for a random sample of 335 treated end-stage renal disease patients who died between 1980 and 1986 in Maryland. On the registry form, patient death was ascribed to 1 of 22 precoded causes by the patient's nephrologist of record. Corresponding death certificates were coded, according to rules of the *International Classification of Diseases*, 9th edition, by a trained observer unaware of the registry report. Agreement was measured by the kappa statistic.

Results. Overall cause of death agreement was poor (31%), and varied by the following categories: renal disease (40% on death certificates vs 0% on registry reports), cardiovascular disease (26% vs 47%), infectious disease (16% vs 22%), cancer (7% vs 5%), and withdrawal from therapy (0% vs 3%). Agreement was higher for transplant recipients than for dialyzed patients.

Conclusions. Death certificates and registry reports yield different descriptions of mortality in end-stage renal disease patients. These sources of information should not be used interchangeably. Improvements to *International Classification of Diseases* coding of renal diseases and the determination of the reliability and validity of the US Renal Data System reporting process are necessary steps in the development of renal disease epidemiology. (*Am J Public Health.* 1993;83:1735–1738)

Introduction

Limitations of mortality statistics were first discussed by Graunt in his *Bills of Mortality*.[1] The cause of death selected depends on the source of information[2]: clinical records, autopsy findings, and death certificates consistently disagree on cause of death.[3–5] The selection is also affected by the disease classification system in use.[2] Standard rules proposed by the *International Classification of Diseases*, 9th edition[6] (ICD-9), are deemed "suboptimal for many biomedical applications,"[7] which explains why disease registries often use customized cause-of-death reporting procedures.

One such registry, the US Renal Data System,[8] collects information, including cause-of-death reports, on patients treated for end-stage renal disease in the Medicare End-Stage Renal Disease Program. Regional registries[9,10] and the national US Renal Data System[11] identify cardiovascular and infectious diseases as the most common causes of death in end-stage renal disease patients. To assess the reliability of this information, we examined the level of agreement between causes of death reported to the Maryland regional registry (now part of the US Renal Data System) and corresponding information derived from death certificates.

Methods

The study sample (n = 378) was drawn at random from 757 patients enrolled in the Maryland Medicare End-Stage Renal Disease Program who died between January 1, 1980, and December 31, 1986.

Matching of Maryland registry information with death certificates, based on name, age, and date of death, was successful for 335 (89%) patients. For 37 patients, linkage was accepted despite a minor mismatch: 1-day difference in date of death (n = 8), middle name used as first name (n = 3), misspelled first name (n = 10), or one-letter substitution in last name (n = 16). The remaining 43 patients (11%) died outside Maryland, or their records contained errors causing a mismatch.

Cause of death was reported to the registry by the patient's attending nephrologist according to 1 of 22 categories specified on the end-stage renal disease death notification form (numbered categories in Table 2). Photocopies of original death certificates were obtained from the Maryland Vital Statistics Office. An investigator (T.V.P.) trained in ICD-9 coding[6,12] selected the underlying cause of death on all death certificates without knowledge of the cause reported to the registry. Coding was repeated 3 months later. Twenty-three inconsistencies between the first and second codings were resolved by formal reference to coding rules. Causes of death derived from death certificates were grouped according to the 22 registry cause-of-death

Thomas V. Perneger is with the Department of Epidemiology, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, MD, and the Institute of Social and Preventive Medicine, University of Geneva, Geneva, Switzerland. Michael J. Klag is with the Department of Medicine, School of Medicine, and the Departments of Epidemiology and Health Policy and Management, School of Hygiene and Public Health, The Johns Hopkins University. Paul K. Whelton is with the Department of Epidemiology, School of Hygiene and Public Health, and the Department of Medicine, School of Medicine, The Johns Hopkins University.

Requests for reprints should be sent to Paul K. Whelton, MD, MSc, Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins Health Institutions, 292 Carnegie, 600 N Wolfe St, Baltimore, MD 21205.

This paper was accepted April 29, 1993.

TABLE 1—Characteristics of 335 Patients with End-Stage Renal Disease Who Died in Maryland between 1990 and 1986

	No.	%
Gender		
Men	192	57
Women	143	43
Race		
White	194	58
Other	141	42
Age at death, y		
< 65	188	56
≥ 65	147	44
Renal transplant		
Yes	29	9
No	306	91
Interval between onset of end-stage renal disease and death, mo		
≤6	90	27
7–12	67	20
13–24	97	29
25–48	58	17
> 48	23	7
Place of death		
Hospital	303	91
Home	32	9
Autopsy		
Yes	47	14
No	288	86

TABLE 2—Causes of Death Reported on Death Certificates and to the Medicare End-Stage Renal Disease Registry and Agreement Between the Sources of Information

Cause of Death (Registry Code)	Death Certificate No.	Death Certificate %	Registry No.	Registry %	Kappa
Pericarditis (1)	...	...	6	2	0.00
Myocardial infarction (2)	15	4	38	11	0.38
Heart disease, other (3)	56	17	82	24	0.26
Cerebrovascular disease (4)	5	1	25	7	0.18
Air embolism (5)	...	...	...	...	...
Pulmonary embolism (6)	2	1	6	2	0.24
Gastrointestinal hemorrhage (7)	4	1	9	3	0.14
Vascular access hemorrhage (8)	...	...	...	...	...
Hemorrhage, other (9)	1	<1	3	1	0.50
Pulmonary infection (10)	12	4	17	5	0.39
Septicemia (11)	24	7	44	13	0.29
Viral hepatitis (12)	...	...	2	1	0.00
Infection, other (13)	16	5	11	3	0.19
Hyperkalemia (14)	...	...	10	3	0.00
Pancreatitis (15)	1	<1	1	<1	1.00
Neoplasia (16)	23	7	16	5	0.76
Withdrawal from therapy (17)	...	...	11	3	0.00
Suicide (18)	1	<1	2	1	0.67
Accident, treatment related (19)	1	<1	...	...	0.00
Accident, other (20)	1	<1	1	<1	1.00
Unknown or unreported (21)	3	1	33	10	0.04
Specified cause, other (22)	21	6	18	5	0.03
Hypertensive renal disease	12	4	...	...	0.00
Hypertensive heart and renal disease	9	3	...	...	0.00
Diabetes, with nephropathy	36	11	...	...	0.00
Diabetes, without nephropathy	7	2	...	...	0.00
Chronic renal failure or end-stage renal disease	82	24	...	...	0.00
Renal disease, other	3	1	...	...	0.00
Cardiovascular disease (1 to 6 and hypertensive heart and renal disease)	87	26	157	47	0.31
Infectious disease (10 to 13)	52	16	74	22	0.44
Kidney disease[a]	133	40	...	...	0.00

[a]Hypertensive renal disease, diabetes with nephropathy, chronic renal failure or end-stage renal disease, and renal disease, other.

categories. Categories for renal diseases and for diabetes were added because no equivalent was found on the registry list (unnumbered categories in Table 2).

Frequency distributions, cross tabulations, agreement between sources, and corresponding kappa statistics were obtained. The kappa statistic represents agreement achieved beyond that expected by chance alone.[13,14] A 20 × 20 cross tabulation (registry codes 5 and 8, corresponding to air embolism and vascular access hemorrhage, were not used) was used to compute overall agreement and the kappa statistic in the total sample and for demographic strata. A 2 × 2 table (specified cause vs other) was used to compute disease-specific agreement and kappa in the total sample only. National Institutes of Health General Clinical Research Centers computerized data management and SAS software were used in conducting analyses.

Results

The study sample was approximately equally divided by gender and race (Table 1). Of the non-Whites, 140 were Black and 1 was Hispanic. The average age at death was 54 years.

Distributions of cause of death showed a striking difference ($\chi^2(19) = 209$, $P < .001$) between the two sources of information (Table 2). In registry reports, the most frequent cause of death was cardiovascular disease (47%), followed by infectious diseases (22%). Renal failure was not reported as an underlying cause of death, except as reflected by the 27 (8%) deaths attributed to pericarditis, hyperkalemia, and withdrawal from therapy. No deaths were attributed to diabetes mellitus. In 1 in 10 patients, the cause of death was reported as unknown or the information was missing.

The most frequent cause of death reported on death certificates was renal disease. Forty percent of deaths were ascribed to specific kidney diseases, chronic renal failure, or end-stage renal disease. Cardiovascular diseases accounted for one in four deaths, and diabetes mellitus accounted for 13% of deaths. No deaths were ascribed to withdrawal from therapy, although it was mentioned as a contributing cause on three death certificates.

The overall agreement was 31%, and the corresponding kappa was .24 (Table 3). Agreement varied little by gender, race, or age; better agreement was achieved among transplanted than among nontransplanted patients. Agreement did not vary meaningfully with increasing interval between initiation of end-stage renal disease therapy and death. Disease-specific agreement was good for cancer, moderate for infectious disease, and poor for cardiovascular disease.

Discussion

This study revealed substantial differences between cause-of-death reports made to two information systems for the

TABLE 3—Agreement on Causes of Death between Death Certificates and Medicare End-Stage Renal Disease Registry Reports

Group	No.	Observed Agreement, %	Kappa
Men			
White	118	34.0	0.27
Black	74	32.4	0.25
Women			
White	76	30.3	0.22
Black	67	23.9	0.17
Age, y			
< 65	188	27.1	0.21
≥ 65	147	35.4	0.28
Transplant			
No	306	28.8	0.21
Yes	29	51.7	0.47
Interval between end-stage renal disease and death, y			
≤ 2	254	32.0	0.25
> 2	81	27.0	0.21
Overall	335	30.8	0.24

same end-stage renal disease patients. Cardiovascular diseases and complications of therapy were prominent in registry reports, while renal diseases were most frequent on standard death certificates. The two sources agreed in fewer than a third of the patients, and the kappa statistic confirmed poor agreement.[14]

The discrepancy could stem from insufficient reliability or validity of either or both reporting systems. Reliability may have been affected by the lack of training of the certifiers and by interrater variability, given that different persons usually provided information to each source. Validity of cause-of-death assessment in patients with renal failure may be particularly problematic. Uremia was recognized as a poorly documented cause of death in 1912.[15] Since then, trends in kidney disease mortality have often been discussed in terms of changes in certification patterns.[16–20] Before widespread availability of end-stage renal disease therapy, death certificates identified renal disease in only 67% of patients dying of uremia.[21] Even when mentioned on a death certificate, a renal disease is coded as the underlying cause of death only about 20% of the time.[22]

Additional problems are inherent to the ICD-9. Today, most patients with end-stage renal disease do not have a "primary" renal disease but suffer from complications of diabetes or hypertension. Nephropathy complicating a systemic disease is usually indicated by the third or fourth digit in the ICD-9 code, which is routinely omitted in summary statistics and may be misstated (for diabetic nephropathy, the ICD-9 code is 250.3, but the code is 250.4 in the ICD-9 Clinical Modification). Moreover, selection of the *underlying* cause of death is inconsistent: for "myocardial infarction due to diabetes with nephropathy," diabetes mellitus is the underlying cause of death, while for "myocardial infarction due to hypertensive renal disease," coronary disease overrules hypertensive disease by a specific coding rule.

Another plausible explanation is that death certificates and registry reports do not measure the same concept of "cause of death." The registry reporting system emphasizes possible indicators of quality of care in the Medicare End-Stage Renal Disease Program, such as withdrawal from therapy, vascular access hemorrhage, or hyperkalemia, which may be poorly reported on standard death certificates. These are mechanisms of death rather than underlying causes. Incidentally, even agreement on mechanism of death can be difficult to achieve: recognition of "social and contextual factors" is incompatible with the registry's emphasis on biological mechanisms.[23]

The registry reporting of cause of death may address legitimate needs of the Medicare End-Stage Renal Disease Program, but it has limited external validity. Notably, the registry list includes neither renal nor systemic diseases as possible causes of death in end-stage renal disease patients. Thus, US Renal Data System mortality statistics cannot be compared with ICD-based mortality reports, and studies using registry data[10] may have little in common with studies using medical records[24] or a combination of sources.[25] Selection of a cause of death in end-stage renal disease patients deserves more attention from renal disease nosologists and US Renal Data System information managers. Adjustments to the Renal Data System database may include coding of multiple cause-of-death data[26] and separate reporting of cause-of-death and quality control information. Clarification of ICD-9 coding for kidney diseases is also warranted. Increased compatibility between the two information systems may greatly increase their usefulness. □

Acknowledgments

Dr Perneger was supported by grants from the Swiss Science Foundation (823B-025121 and 32-32609.91) and the Schmidheiny Foundation. Dr Klag is an established investigator of the American Heart Association. Dr Whelton was supported by General Clinical Research Centers grant 5MOIRR00722 from the National Center for Research Resources, National Institutes of Health.

This research was presented at the 24th Annual Meeting of the American Society of Nephrology, November 1991.

Computational assistance was received from the General Clinical Research Centers Computing Center (supported by National Institutes of Health grant R00035).

References

1. Graunt J. *Natural and Political Observations Mentioned in a Following Index, and Made upon the Bills of Mortality.* Baltimore, Md: The Johns Hopkins Press; 1939. Reprint.
2. Lilienfeld AM, Lilienfeld DE. Mortality statistics. In: *Foundations of Epidemiology.* 2nd ed. New York, NY: Oxford University Press; 1980:66–83.
3. Battle RM, Pathak D, Humble CG, et al. Factors influencing discrepancies between premortem and postmortem diagnoses. *JAMA.* 1987;258:339–344.
4. Kircher T, Nelson J, Burdo H. The autopsy as a measure of accuracy of the death certificate. *N Engl J Med.* 1985;313:1267–1273.
5. Gittelsohn A, Senning J. Studies on the reliability of vital and health records: I. Comparison of cause of death and hospital record diagnoses. *Am J Public Health.* 1979;69:680–689.
6. *Manual of the International Classification of Diseases, Injuries and Cause of Death, 1975, 9th Revision.* Geneva, Switzerland: World Health Organization; 1977.
7. Carter JR. The problematic death certificate. *N Engl J Med.* 1985;313:1285–1286.
8. United States Renal Data System. Description of the USRDS and its data base. *Am J Kidney Dis.* 1991;18(suppl 2):17–20.
9. Port FK. Mortality and causes of death in patients with end-stage renal failure. *Am J Kidney Dis.* 1990;15:215–217.
10. Mailloux LU, Bellucci AG, Wilkes BM, et al. Mortality in dialysis patients: analysis of the causes of death. *Am J Kidney Dis.* 1991; 18:326–335.
11. US Renal Data System. *USRDS 1991 Annual Data Report.* Bethesda, Md: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 1991.
12. National Center for Health Statistics. *Instructions Manual Part 2a. Instructions for Classifying the Underlying Cause of Death, 1990.* Washington, DC: US Dept of Health and Human Services; 1990.
13. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: Wiley; 1981.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
15. Cabot RC. Diagnostic pitfalls identified during a study of three thousand autopsies. *JAMA.* 1912;59:2295–2298.
16. Kessner DM, Florey CV. Mortality trends for acute and chronic nephritis and infec-

tions of the kidney. *Lancet.* 1967;2:979–982.
17. Waters WE. Trends in mortality from nephritis and infections of the kidney in England and Wales. *Lancet.* 1968;1:241–243.
18. Paffenbarger RS, Milling RN, Poe ND, Krueger DE. Trends in death rates from hypertensive disease in Memphis, Tennessee, 1920–1969. *J Chronic Dis.* 1966;19:847–856.
19. Varis J, Oksa H, Pasternack A. Mortality from renal diseases in Finland. *Scand J Urol Nephrol.* 1988;22:147–153.
20. Hansen H, Susser M. Historic trends in deaths from chronic kidney disease in the United States and Britain. *Am J Epidemiol.* 1971;93:413–424.
21. Modan B, Bott-Kanner G, Barnoach N, Leslau V, Eliahou HE. Chronic renal disease in Israel. Validity of death certificates. *Isr J Med Sci.* 1971;7:1550–1553.
22. Seigel D, Markush RE. Prevalence at death II. Methodological considerations for use in mortality studies. *Am J Public Health.* 1968;58:772–776.
23. Plough AL, Salem S. Social and contextual factors in the analysis of mortality in end-stage renal disease patients: implications for health policy. *Am J Public Health.* 1982;72:1293–1295.
24. Washer GF, Schroter GPJ, Starzl TE, Weil R. Causes of death after kidney transplantation. *JAMA.* 1983;250:49–54.
25. Hill MN, Grossman RA, Feldman HI, Hurwitz S, Dafoe DC. Changes in causes of death after renal transplantation, 1966 to 1987. *Am J Kidney Dis.* 1991;17:512–518.
26. Israel RA, Rosenberg HM, Curtin LR. Analytical potential for multiple cause-of-death data. *Am J Epidemiol.* 1986;124:161–179.

Public Health Briefs

Infectious Diseases and Mortality among US Nursing Home Residents

Consuelo Beck-Sague, MD, Shailen Banerjee, PhD, and William R. Jarvis, MD

ABSTRACT

Data collected in the 1985 National Nursing Home Survey were analyzed to identify risk factors for infections and mortality and to explore their relationship in US nursing homes. An infection was recorded in 166 609 (14%) discharges. Risk of pneumonia was found to be higher among bedbound patients (54.5 vs 13.1 per 100 discharges); urinary tract and other infections were most frequent among residents with indwelling catheters (6.6 vs 1.0 per 100 discharges). Residents with pneumonia were more likely than those with other infections to die (35% vs 28%), or be discharged to hospitals if alive (94% vs 66%). Thus, immobility and catheterization were associated with infections in US nursing homes, and pneumonia was found to contribute to mortality. (*Am J Public Health.* 1993;83:1739–1742)

Introduction

Of the 2.1 million persons over age 85 in the United States, 22% reside in nursing homes.[1] The population of persons aged 85 years and older will exceed 4 million by the year 2000.[2] Risk of infection in nursing homes ranges from 3.3% to more than 15%.[3–5]

Infections and their relationship to mortality in nursing homes have been studied.[5–9] This is the first study that uses a population-based national survey to identify risk factors for infections and mortality and to explore their relationship in US nursing homes.

Methods

Collection of Data

Data for this study were taken from the 1985 National Nursing Home Survey conducted by the National Center for Health Statistics (NCHS). The survey included all types of nursing and related-care homes with at least three beds.[10] Sample data were collected from 20 479 US nursing homes by applying a stratified two-stage probability design; methods have been described previously.[10] Data collected included measures of the functions of residents, and admission and discharge diagnoses, coded using the *International Classification of Diseases,* 9th revision, Clinical Modification.[11]

Case Definition

Discharges were defined as cases of infection if the survey interviewer recorded an infection at discharge. These diagnoses were grouped in coding as pneumonia, urinary tract infection, and infectious and parasitic diseases.

Statistical Methods

Rates of infectious diseases were calculated for discharged residents by selected resident characteristics. Because some residents had infections diagnosed on admission, residents with the same infection at admission and discharge were excluded from the analysis so that the risk associated with nursing home residence could be assessed. Weights provided by the NCHS were used to obtain estimates for national averages.[10] Standard errors for relative risks were obtained by using constants obtained from the NCHS.[10,12,13] Rate ratios and their 95% confidence intervals were calculated as estimates of relative risk.[12] Significance of relative risks was tested by calculating standard normal variates and the corresponding probabilities. Duration of stay was estimated for groups of discharged residents and compared by calculating chi-squares and corresponding probabilities from the contingency tables generated.[14] Infections among discharge diagnoses were compared by month of discharge to identify seasonal variation in the incidence of infections and mortality. Rates of discharge to health care facilities versus residential settings and rates of fatality were also calculated and compared.

Results

Study Population

We reviewed information on 1 220 816 residents (projected from a sample of 6017

The authors are with the Hospital Infections Program, Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Ga.

Requests for reprints should be sent to Consuelo Beck-Sague, MD, Hospital Infections Program, Bldg 3 B-49A Stop A07, Centers for Disease Control and Prevention, Atlanta, GA 30333.

This paper was accepted May 6, 1993.

TABLE 1—Rates of Urinary Tract and Other Bacterial, Viral, and Parasitic Infections among Discharges from US Nursing Homes, by Selected Resident Characteristics

	Sample Discharges (n = 5622)[a]	Population Estimate of Discharges (n = 1 130 974)	Infection Rate[b] (n = 33 050)	Relative Risk	95% Confidence Interval
Indwelling catheters					
Catheterized	481	113 920	6.6	6.5	3.2, 13.4
Continent for urine	2361	488 145	1.0		
Incontinence, stool					
Incontinent	2485	485 343	4.5	3.3	1.9, 5.5
Continent	2667	556 096	1.4		
Ambulation					
Bedfast	2048	384 604	4.1	2.0	1.2, 3.1
Not bedfast	3369	707 400	2.1		
Incontinence, urinary					
Incontinent, not catheterized	2361	466 079	3.9	3.8	2.0, 7.2
Continent	2429	488 145	1.0		
Sex					
Male	1986	417 487	2.8	3.2	0.7, 1.8
Female	3636	713 487	1.2		

Source. 1985 National Nursing Home Survey.
[a]Number in sample, excluding discharges with same infection(s) at admission and discharge and those in whom information was missing.
[b]Infections per 100 discharged residents (n = projected number of discharges with urinary tract and other bacterial, viral, and parasitic infections).

TABLE 2—Rates of Pneumonia among Discharges from US Nursing Homes, by Selected Resident Characteristics

	Sample Discharges (n = 5803)[a]	Population Estimate of Discharges (n = 1 171 154)	Pneumonia Rate[b] (n = 60 334)	Relative Risk	95% Confidence Interval
Indwelling catheters					
Catheterized	520	125 943	4.6	1.8	0.9, 3.3
Continent for urine	2398	497 602	2.6		
Incontinence, stool					
Incontinent	2596	509 909	7.9	3.0	2.0, 4.3
Continent	2729	569 266	2.6		
Ambulation					
Bedfast	2124	402 434	8.2	2.5	1.8, 3.4
Not bedfast	3464	726 198	3.3		
Incontinence, urinary					
Incontinent, not catheterized	2398	484 496	7.8	3.0	2.0, 4.5
Continent	2531	497 602	2.6		
Sex					
Male	2026	428 187	6.2	1.4	1.0, 1.9
Female	3775	742 967	4.6		

Source. 1985 National Nursing Home Survey.
[a]Number in sample, excluding discharges with pneumonia at both admission and discharge and those in whom information was missing.
[b]Number with pneumonia per 100 discharged residents (n = projected number of discharges with pneumonia).

residents) discharged from 19 068 US nursing homes (projected from 1078 homes). These residents ranged in age at discharge from 2 to 109 years (median = 82 years) and in duration of residence from less than 1 month to 53 years (median = 4.5 months). Additionally, 62.8% were female.

Infectious Diseases in Discharged Residents

Infections were included in the diagnoses at discharge of 166 610 (14%) residents. In 82 192 (49%), pneumonia was among the discharge diagnoses; in 56 666 (34%), urinary tract infections were included; and in 27 752 (17%), bacterial, viral, and parasitic infections were cited. Among 84 695 (51%) residents with primary diagnoses of infections at discharge, pneumonia (64%), urinary tract infection (19%), or another infection (17%) was the primary discharge diagnosis.

In 33 050 discharged residents with urinary tract infections and bacterial, viral, and parasitic infections, these infections were not included in the admission diagnoses, which suggests that they were acquired in the nursing home. Rates of infection were significantly higher among residents who were bedbound (4.1 vs 2.1) or incontinent for stool (4.5 vs 1.4) or urine (3.9 vs 1.0) (Table 1). The highest risk was associated with indwelling catheterization (6.6 vs 1.0).

In 60 334 residents with a discharge diagnosis of pneumonia, pneumonia was not included in the admission diagnoses, which suggests again that, in these residents, the infection was acquired in the nursing home. Rates of pneumonia at discharge were highest among residents who were bedbound (8.2 vs 3.3) or incontinent for stool (7.9 vs 2.6) or urine (7.8 vs 2.6) (Table 2). Indwelling catheterization was not as strongly predictive for pneumonia as it was for other infections, and the risk was not statistically significant. Prior stay in the nursing home was longer for residents with pneumonia (median = 6.4 months) than for other residents (median = 4.3 months, $P < .001$).

Seasonal variation was apparent among residents who had pneumonia at discharge (Figure 1). Almost 50% were discharged from January through April. Peak pneumonia and mortality associated with pneumonia coincided with the peak incidence of influenza cases in the United States during 1985 and 1986.[15] No seasonal variation was seen among discharges with infections other than pneumonia (Figure 1).

Fatality and Discharge Outcomes

Fatality rates were highest among residents who were bedbound (54.5 vs 13.1), incontinent (46.7 vs 11.0), or catheterized prior to discharge (4.3 vs 9.7), or who had pneumonia (35.3 vs 28.1) (Table 3). Among residents discharged alive, those whose di-

agnoses included pneumonia were more likely to be discharged to a health care facility than other residents (54 507/58 248 [94%] vs 535 937/817 656 [66%], $P < .001$), who were more likely to transfer to residential settings. Residents whose diagnoses included urinary tract infections and other infections excluding pneumonia were also more likely to be transferred to health care facilities (50 697/60 447 [84%] vs 485 240/757 209 [64%], $P < .001$).

Discussion

Pneumonia appears to be associated with considerable morbidity and mortality in US nursing homes. The striking seasonal variation of pneumonia as a discharge diagnosis in nursing homes suggests an association with influenza activity. It has been recognized that more than 80% of influenza-related deaths occur in the elderly.[16,17] Increases in immunization levels in nursing homes from 36% in the 1979/80 season to 71% in the 1985/86 season were reported, but the influenza epidemic of 1985/86 was characterized by considerable pneumonia mortality among elderly institutionalized persons.[16–19] A low immunization level is associated with the occurrence of influenza outbreaks in US nursing homes.[19]

Factors related to infections and death in nursing home patients include nonambulatory status. Aspiration and other outcomes associated with immobility have been previously identified as risk factors for urinary tract infection, pneumonia, and death.[4,9,20–23] Rehabilitation of elderly nursing home residents is controversial because the need for restoring mobility may be poorly understood, but mobility may reduce incontinence, infections, and mortality.[24–26] Delay in evaluating the signs of infection and instituting treatment may also contribute to the high fatality rate of residents with pneumonia.[20,27]

Assessment of infection control procedures in nursing homes has identified infrequent use of proven methods to reduce risk of urinary tract infections.[28] The sealing of catheters and use of prophylactic antimicrobials should be explored.[29]

Some aspects of the survey and study design limit the applicability of these findings. No information was available to assess the preventability of risk factors or to evaluate infection control or immunization programs. Exclusion of residents with the same infection at admission and discharge may have also excluded some residents for whom the discharge infection was a completely new event. Those with infections that were acquired in the nursing home and resolved before discharge were excluded. Finally, population estimates for some factors are based on small (fewer than 1000 discharges) samples. Nevertheless, this study suggests that

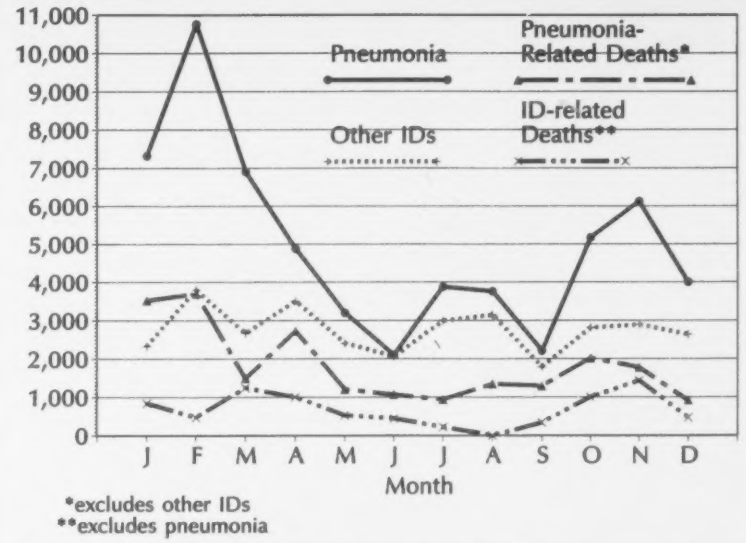


FIGURE 1—Distribution of discharge diagnoses of pneumonia, pneumonia-related deaths, other infectious diseases (IDs), and deaths due to other IDs, by month, 1985 National Nursing Home Survey.

TABLE 3—Fatality Rate among Discharges from US Nursing Homes, by Discharge Diagnosis of Infectious Diseases and Other Resident Characteristics

	Sample Discharges (n = 6000)[a]	Population Estimate of Discharges (n = 1 218 956)	Fatality Rate[b] (n = 343 051)	Relative Risk	95% Confidence Interval
Indwelling catheters					
Catheterized	546	132 833	43.6	4.5	3.6, 5.5
Continent for urine	2455	510 323	9.7		
Incontinence, stool					
Incontinent	2711	536 446	46.7	4.2	3.6, 5.0
Continent	2803	588 527	11.0		
Ambulation					
Bedfast	2222	425 329	54.5	4.1	3.6, 4.7
Not bedfast	3662	752 586	13.1		
Incontinence, urinary					
Incontinent, not catheterized	2638	510 728	42.6	4.4	3.6, 5.3
Continent	2455	510 323	9.7		
Pneumonia					
Diagnosed	395	77 459	35.3	1.3	1.0, 1.5
No pneumonia	5201	1 052 444	28.1		
Infection, excluding pneumonia					
Diagnosed	350	78 599	23.1	0.8	0.6, 1.1
No infection	5201	1 052 444	28.1		
Sex					
Male	2125	453 296	28.5	1.0	0.9, 1.2
Female	3875	765 660	27.9		

Source. 1985 National Nursing Home Survey.
[a]Number in sample, excluding discharges in whom information was not available.
[b]Number dead at discharge per 100 discharged residents (n = projected number of residents dead at discharge).

preventing infections acquired in the nursing home may reduce the risk of nursing home stays. □

Acknowledgment

The authors gratefully acknowledge the assistance of Esther Hing of the Long-Term Care Statistics Branch, National Center for Health Statistics, who reviewed the manuscript and provided technical advice in the data analysis.

References

1. Van Nostrand J, Zappola A, Hing E, et al. The National Nursing Home Survey: 1977 summary. *Vital Health Stat [13].* 1979;43: 28–29. DHHS publication PHS 79-1794.
2. Smith PW. Infections in long-term care facilities. *Infect Control.* 1985;6:435–436. Editorial.
3. Smith PW. Infection control in nursing homes. *JAMA.* 1985;254:2951–2952. Editorial.
4. Setia U, Serventi I, Lorenz P. Nosocomial infections among patients in a long-term care facility: spectrum, prevalence and risk factors. *Am J Infect Control.* 1985;13:57–62.
5. Smith IM. Infectious diseases of the geriatric patients. In: Smith PW, ed. *Infection Control in Long-Term Care Facilities.* New York, NY: John Wiley and Sons; 1984:33–39.
6. Ruben FL, Norden CW, Heisler B, Korica Y. An outbreak of *Streptococcus pyogenes* infections in a nursing home. *Ann Intern Med.* 1984;101:494–496.
7. Franson TR, Duthie EH, Cooper JE, et al. Prevalence survey of infections and their predisposing factors at a hospital-based nursing home care unit. *J Am Geriatr Soc.* 1986;34:95–100.
8. Scheckler WE, Peterson RN. Infections and infection control among residents of eight rural Wisconsin nursing homes. *Arch Intern Med.* 1986;146:1981–1984.
9. Mehr DR, Foxman B, Colombo P. Risk factors for mortality from lower respiratory infections in nursing home patients. *J Fam Pract.* 1992;34:585–591.
10. Hing E, Sekcsenski E, Strahan G. National Nursing Home Survey: 1985 Summary for United States. *Vital Health Stat [13].* 1989; 97. DHHS publication PHS 89–1758.
11. *International Classification of Diseases,* 9th revision, Clinical Modification. Vol 1. Hyattsville, Md: National Center for Health Statistics, 1984. National Committee on Vital and Health Statistics; DHHS publication PHS 80–1260.
12. Kleinbaum D, Kupper L, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods.* Belmont, Calif: Lifetime Learning Publications; 1982.
13. *SAS User's Guide: Basics, Version 5 Edition.* Cary, NC: SAS Institute Inc; 1985.
14. Conover WJ. *Practical Nonparametric Statistics.* New York, NY: John Wiley and Sons; 1980.
15. Centers for Disease Control. Influenza 1985–1986 seasons. *MMWR.* 1986;35:470, 475–479.
16. Williams WW, Hickson MA, Kane MA, et al. Immunization policies and vaccine coverage among adults: the risk for missed opportunities. *Ann Intern Med.* 1988;108: 616–625.
17. Arden NH, Patriarca PA, Kendal AP. Experiences in the use and efficacy of influenza vaccine in nursing homes. In: Kendal A, Patriarca P, eds. *Options for the Control of Influenza.* New York, NY: Alan R. Liss; 1986:155–168.
18. Centers for Disease Control. State and local influenza immunization program activities. *MMWR.* 1988;37:705–707.
19. Patriarca PA, Weber JA, Parker RA, et al. Risk factors for outbreaks of influenza in nursing homes—a case-control study. *Am J Epidemiol.* 1986;124:114–119.
20. Marrie TJ, Durant H, Kwan C. Nursing home-acquired pneumonia: a case-control study. *J Am Geriatr Soc.* 1986;34:697–702.
21. Irvine P, Van Buren N, Crossley K. Causes for hospitalization of nursing home residents. *J Am Geriatr Soc.* 1984;32:103.
22. Alvarez S, Shell CG, Woolley TW, Berk SL, Smith JK. Nosocomial infections in long-term facilities. *J Gerontol.* 1988;43:M9–17.
23. Powers JS, Billings FT, Behrendt D, Burger MC. Antecedent factors in urinary tract infections among nursing home patients. *South Med J.* 1988;81:734–735.
24. Kane DK. Assessing the outcomes of nursing home patients. *J Gerontol.* 1983;38: 385–390.
25. Parry EA. Physical rehabilitation of the old, old patient. *J Am Geriatr Soc.* 1983; 31:482–486.
26. Hesse WK. Attitudinal stumbling blocks to geriatric rehabilitation. *J Am Geriatr Soc.* 1984;32:747–750.
27. Franson TW, Schicker JM, Le Clair SM, et al. Documentation and evaluation of fevers in hospital-based and community-based nursing homes. *Infect Control Hosp Epidemiol.* 1988;9:447–450.
28. Khabbaz RF, Tenney JH. Infection control in Maryland nursing homes. *Infect Control Hosp Epidemiol.* 1988;9:159–162.
29. Platt R, Polk BF, Murdock B, Rosner B. Prevention of urinary tract infection: a cost-benefit analysis. *Infect Control Hosp Epidemiol.* 1989;10:60–64.

Nemesis Revisited: Tuberculosis Infection in a New York City Men's Shelter

ABSTRACT

In November 1990, a screening was conducted to determine the point prevalence of tuberculosis infection in a volunteer sample of homeless men (n = 161) living in a congregate shelter in New York City. Of those for whom we have results (n = 134), 79% were positive for tuberculosis. The mean length of shelter stay from date of shelter entry was 31.8 months and was significantly associated with the tuberculosis infection rate. The findings suggest that crowded living conditions and the presence of a stable resident pool in crowded congregate shelters may be associated with transmission of tuberculosis infection. (*Am J Public Health.* 1993;83: 1743–1745

Eugene A. Paul, MD, MPH, Sharon M. Lebowitz, MPH, Robert E. Moore, DrPH, Christina W. Hoven, DrPH, Brenda A. Bennett, MD, MPH, and Abigail Chen

Introduction

Nationally, the number of new cases of tuberculosis increased by 5% in 1989 and by 10% in 1990. The increase has been largely among 25- to 44-year-old men. This demographic finding coincides with—although it is not necessarily causally related to—the epidemic of acquired immunodeficiency syndrome (AIDS) and the increase in homelessness. In New York City alone, 3520 new tuberculosis cases were reported in 1990, representing a 38% increase over the number of new cases in 1989. This incidence level in New York City constitutes approximately 15% of all US cases reported in 1990.[1]

The prevalence of asymptomatic tuberculosis infection among the homeless has been reported to be as high as 22% to 50%.[2,3] In an immunocompetent population, it is estimated that only 25% of those exposed to tuberculosis become infected and that 10% of this subgroup go on to develop clinical disease.[4] However, the homeless are at greater risk for immune suppression because of poor nutrition, inadequate rest, and intercurrent medical illness. Among homeless men, 15% to 25% have also been found to have a lifetime history of a psychotic disorder, and a majority have histories of serious substance use.[5]

In this paper, we report on the prevalence of tuberculosis infection and selected risk factors for infection in a volunteer sample (n = 161) of homeless men who reside in a large congregate shelter (n = 660). We were particularly interested in exploring whether there might be a significant association between the length of stay in the shelter system and the development of tuberculosis infection, given that contagious spread is facilitated by living in overcrowded quarters.[6–8]

Methods

The Fort Washington Men's Shelter is located within a large armory administered by New York City's Human Resources Administration. The Shelter Medical Clinic, operated by Harlem Hospital Center, conducted the screening with the cooperation of the Harlem Hospital AIDS Treatment Group and the New York City Department of Health.

Approximately 660 men on a given night sleep in this shelter (see photo). Beds are separated by approximately 18 to 20 inches and are arranged on a large open floor measuring 55 000 square feet.

The drill floor (sleeping area) was selected as the screening site. Coffee and a snack were offered as an incentive for participation. Notices were posted throughout the shelter several days before the screening, and caseworkers who were in frequent contact with the men were encouraged to refer them to the screening. On November 26, 1990, 161 men volunteered and were evaluated for tuberculosis infection. Brief semistructured interviews were conducted. The date of the first shelter stay, which is entered into the Shelter Care Information Management System and signifies the date an individual was officially registered by a caseworker anywhere in the New York City shelter system, was used to approximate length of shelter stay.

A self-reported history of tuberculosis or a positive skin test in similar populations has been found to represent a valid assessment of tuberculosis status.[9–12] Therefore, all residents who did not have either a history of tubercu-

Eugene A. Paul and Brenda A. Bennett are with the Harlem Hospital Center/College of Physicians and Surgeons, Columbia University, New York, NY. Sharon M. Lebowitz is with the Health and Hospitals Corporation's Office of Homelessness Management, New York, NY. Robert E. Moore, Christina W. Hoven, and Abigail Chen are with the New York State Psychiatric Institute, New York, NY.

Requests for reprints should be sent to Eugene A. Paul, MD, MPH, Health and Hospitals Corporation, Harlem Hospital Center, Department of Ambulatory Care, 506 Lenox Ave, New York, NY 10037.

This paper was accepted June 22, 1993.

Fort Washington men's shelter, drill floor, New York, NY, 1990. ©1990 Patrick Ashley.

TABLE 1—History of Tuberculosis and Skin Test Reactions among Homeless Men in a New York City Shelter, 1990, by Age and Ethnicity (n = 134)[a]

	PPD Negative (n = 29)		PPD Positive by History (n = 33)		PPD Positive (n = 72)	
	n	%	n	%	n	%
Ethnicity						
African-American	22	75.9	22	66.7	60	83.3
Hispanic	7	24.1	11	33.3	12	16.7
Age, y						
20–29	9	31.0	9	27.3	27	37.5
30–39	16	55.2	15	45.4	25	34.7
40–49	4	13.8	6	18.2	14	19.4
50–61	0	0.0	3	9.1	6	8.3

Note. PPD = purified protein derivative.
[a]Excludes those who did not return for a reading (n = 24) or who were anergic (n = 3).

losis or a reactive skin test received (Aplisol) 5 tuberculin (0.1 mL) units of purified protein derivative injected intradermally in the volar aspect of the right forearm. Each resident also received anergy testing with 0.1 mL of 1:1000 strength Candida (in the left forearm). All tests were read within 48 to 96 hours and interpreted according to the American Thoracic Society's recommendations.[13] According to recommendations of the Centers for Disease Control and Prevention (CDC), a purified protein derivative reaction of 10 mm or greater of induration and a Candida reaction of greater than 2 mm of induration were interpreted as positive. Reactions of less than 10 mm were not recorded.

Results

Slightly more than three-fourths (77%) of the total sample (n = 161) identified themselves as African-American, 22% were Hispanic, and one respondent did not reveal his ethnicity. The mean age of the sample was 34.5 (SD = 8.6; range = 20 to 61 years). The mean length of time from date of shelter entry was found to be 31.8 months (SD = 18.6). Time since date of entry into the shelter system ranged from 0 months (2 weeks or less) to 8 years; the median was 29 months. Sixty-four percent of the total sample were defined as long-term users of the New York City shelter system, the criterion being a date of first shelter stay of more than 2 years ago.

TABLE 2—Risk of Tuberculosis Infection among Homeless Men in a New York City Shelter, 1990, by Age, Ethnicity, and Length of Shelter Stay

Predictor	Adjusted Odds Ratio[a]	95% Confidence Interval
Age, y	1.04	0.99, 1.11
African-American vs Hispanic	1.11	0.40, 3.13
Time from date of shelter entry, mo	1.02	1.00, 1.05

[a]Based on results of logistic regression model and adjusted for the other variables in the table.

TABLE 3—Predicted Probability of Tuberculosis Infection among Homeless Men in a New York City Shelter, 1990, by Time from Date of Shelter Entry[a]

	Months					
	0	6	12	18	24	36
Probability	.10	.12	.14	.16	.18	.24

[a]Based on results of logistic regression model.

Those men who were purified protein derivative positive by history and therefore were not tested at the screening represented 21% (n = 33) of the sample (see Table 1). One hundred twenty-eight participants were tested and 81.3% (n = 104) returned for a reading. Of those who returned, 69% (n = 72) were positive; three men failed to react to both purified protein derivative and Candida control. The mean induration value in millimeters for those who returned and tested positive was 18.8 (SD = 4.9, range 10 to 30). The breakdown in 5-mm increments was as follows: 10 to 14 mm: 16%; 15 to 19 mm: 35%; 20 to 24 mm: 35%; and 25 to 30 mm: 14%; only 21% were purified protein derivative negative.

The prevalence rate of any tuberculosis infection, including those participants with an induration of 10 mm or greater and those with past history, excluding the nonreturns, was 79%.

In a logistic regression predicting tuberculosis infection (based on past history or an induration of 10 mm or greater), time from date of entry into the shelter system was significantly associated ($P < .05$) with risk (Table 2). For example, control-

ling for age and ethnicity, the predicted probability of tuberculosis infection at 36 months from date of entry into the shelter system was .24 compared with only .14 at 12 months (Table 3).

Discussion

The most remarkable finding of this study was the extremely high prevalence, past or present, of tuberculosis infection (79%) among those who volunteered for screening. Also, it is notable that we, like others,[8] found that the length of shelter stay was significantly associated with the tuberculosis infection rate. These two findings support the proposition that crowded living conditions in congregate shelters are associated with the spread of tuberculosis infection.

Our study has some important limitations. First, the sample was self-selected and so potential selection bias was introduced. For example, caseworkers may have been more likely to refer medically ill clients to the screening or to have encouraged men at risk of tuberculosis infection (e.g., known intravenous drug users) to participate. However, we found no evidence of this bias.

Second, because an individual may not have stayed in the shelter continuously from entry date to date of screening, the length of shelter stay should be considered a proxy. Third, we had incomplete control of confounders, such as alcohol, drug use, and history of sexually transmitted disease and human immunodeficiency virus (HIV). Finally, only an induration of 10 mm or greater was interpreted as a positive test because, at the time of this screening, 10 mm was still what the CDC recommended for purified protein derivative test positivity in homeless individuals. Since the HIV status of the participants was unknown, 10 mm was used instead of 5 mm to capture all who would unquestionably be considered positive for purified protein derivative in this population. Given that this population is at high risk for HIV disease, one could speculate that the prevalence of purified protein derivative positivity was actually underrepresented since there may have been several men with an induration of between 5 and 9 mm. Based upon current CDC recommendations,[14] members of this group would be candidates for prophylaxis. Therefore, the rate of reported tuberculosis infection and implications for intervention may be even greater.

Local health departments can enable agencies to provide relatively safe tuberculosis-controlled living environments by mandating periodic tuberculosis screening and a follow-up chest x-ray and sputum analysis. Without an effective plan for follow-up, however, screening efforts will serve little purpose. Departments of Health must ensure that patients remain under medical supervision until treatment is completed. The recent epidemic of deaths due to multidrug-resistant tuberculosis strains reported in the New York metropolitan region[15] underscores the fatal consequences of episodic treatment of persons, especially those with immune deficiency. One might suspect that the costs of attempting to treat multidrug-resistant tuberculosis in terms of prolonged, recurrent hospitalization and possible surgery would far exceed those of using directly observed therapy, even with monetary incentives or the equivalent. However, for selected populations, especially among the disenfranchised, supervised therapy with incentives should be seriously considered. The eradication of tuberculosis will come about by social measures as well as by the development of better drugs and vaccines. □

References

1. New York City Department of Health. Tuberculosis, New York City, the 1990 experience. *City Health Info.* 1991;10:1–6.
2. Schieffelbein CW, Snider DE. Tuberculosis control among homeless populations. *Arch Intern Med.* 1988;148:1843–1846.
3. Slutkin G. Management of tuberculosis in urban homeless indigents. *Public Health Rep.* 1986;101:481–494.
4. Dowling PT. Return of tuberculosis: screening and preventive therapy. *Am Fam Physician.* 1991;43:457–467.
5. Susser E, Struening EL, Conover S. Psychiatric problems of homeless men. *Arch Gen Psychiatry.* 1989;46(September):845–850.
6. Riley RL. Airborne infection. *Am J Med.* 1974;57:466–475.
7. Riley RL, Nardell EA. Clearing the air: the theory and application of ultraviolet air disinfection. *Am Rev Respir Dis.* 1989;139:1286–1294.
8. Nolan CM, Elarth AM, Barr H, Saeed AM, Riser DR. An outbreak of tuberculosis in a shelter for homeless men: a description of its evolution and control. *Am Rev Resp Dis.* 1991;143:257–261.
9. McAdam JM, Brickner PW, Scharer LL, Crocco JA, Duff AE. The spectrum of tuberculosis in a New York City men's shelter clinic (1982–1988). *Chest.* 1990;97:798–805.
10. Torres RA, Mani S, Altholz J, Brickner PW. Human immunodeficiency virus infection among homeless men in a New York City shelter: association with *Mycobacterium tuberculosis* infection. *Arch Intern Med.* 1990;150:2030–2033.
11. Brickner PW, Scharer LL, Conanan B, Elvy A, Sararese M. *Health Care of Homeless People.* New York, NY: Springer; 1985.
12. Brickner PW, Scharer LL, Conanan B, Sararese M, Scanlon BC. *Under the Safety Net.* New York, NY: WW Norton Co; 1990.
13. American Thoracic Society diagnostic standards and classification of tuberculosis. *Am Rev Respir Dis.* 1990;142:725–735.
14. Centers for Disease Control. Purified protein derivative (PPD)-tuberculin anergy and HIV infection: guidelines for anergy testing and management of anergic persons at risk of tuberculosis. *MMWR.* 1991;40:27–32.
15. Centers for Disease Control. Transmission of multidrug resistant tuberculosis among immunocompromised persons in a correctional system—New York, 1991. *MMWR.* 1992;41:507.

The School-Based Immunization Survey: An Inexpensive Tool for Measuring Vaccine Coverage

ABSTRACT

A school-based immunization survey was conducted among the 36 Rochester, NY, elementary schools (n = 5584 children) to determine (1) the vaccination rates at 2 years of age by type of primary care provider and (2) the accuracy of school immunization records (by comparing them with medical charts for children attending hospital-based clinics). These rates varied by provider type from 58% to 86% and were all below the national goal of 90%. In comparison with medical chart review, the school data had error rates of 15%; however, these errors occurred in both directions and were thus unbiased. School-based surveys include children who lack connections to the primary care system. With minimal effort these surveys can help identify populations in need of intervention. (*Am J Public Health.* 1993;83:1749–1751)

Lance E. Rodewald, MD, Klaus J. Roghmann, PhD, Peter G. Szilagyi, MD, MPH, Nancy L. Winter, MPA, James R. Campbell, MD, and Sharon G. Humiston, MD, MPH

Introduction

School-based immunization surveys are frequently used to assess vaccination coverage for communities. Alternatives to such surveys have limitations. For example, random samples of children seldom provide enough data to determine vaccination levels of small areas such as neighborhoods served by schools. Immunization surveys of medical charts in providers' offices cannot assess vaccination levels of groups of interacting children because these children receive care from different providers. Schools, however, cover complete, interacting populations and keep comprehensive immunization histories of their students, even if vaccines are administered by multiple providers. The school-based immunization survey provides a unique and simple method to study vaccination rates of interacting populations.

In spite of these advantages, the methodology for school-based immunization surveys is not well described. Results of a few published surveys indicate low vaccination rates,[1–3] implying that many neighborhoods are at risk of outbreaks. However, little has been written on the methods of procedure, cost, strengths, and weaknesses of the school-based survey.[4] We report on an immunization survey of the Rochester, NY, City School District and discuss the methodological issues that pertain to data quality assurance.

Methods

Permission was obtained to analyze the immunization records of children beginning kindergarten in 1991 or 1992 in the Rochester City School District from the city school district, the Monroe County Department of Health, and the University of Rochester Research Subjects Review Board. All 36 city elementary schools participated in the project.

The following data were recorded: school attended, initials of the child, birth date, zip code, and primary care provider. The vaccines recorded were diphtheria, tetanus, and pertussis (DTP); oral polio vaccine (OPV); measles, mumps, and rubella (MMR); and *Hemophilus influenzae* type b (Hib). Inactivated poliovirus vaccine was recorded as OPV. Programs were written to re-create individual immunization histories into temporal sequence, calculate age-specific vaccination rates, and analyze these rates.

The following rates were calculated for each child: (1) those up to date for *all* vaccines at bimonthly intervals between 10 months and 72 months and (2) those up to date for *individual* vaccines by the same ages. The American Academy of Pediatrics *Red Book* recommendations were used to determine standard vaccination ages.[5] Specific algorithms, programs, and record abstraction forms are available from the authors.

The quality of the data was tested by two methods. First, inconsistent dates were identified and checked against the data entry form or against the original school record. Second, a random sample of charts that mentioned either of the two hospital-based pediatric continuity clinics as the student's primary provider was obtained. The charts were abstracted by using the same data collection forms; correlations between the school and provider records were determined.

Results

There were 2654 records abstracted for the 1990 to 1991 kindergartners and 2939 records abstracted for the 1991 to 1992 kindergartners, representing two birth cohorts of city school district children.

The authors are with the Department of Pediatrics, University of Rochester, Rochester, NY.

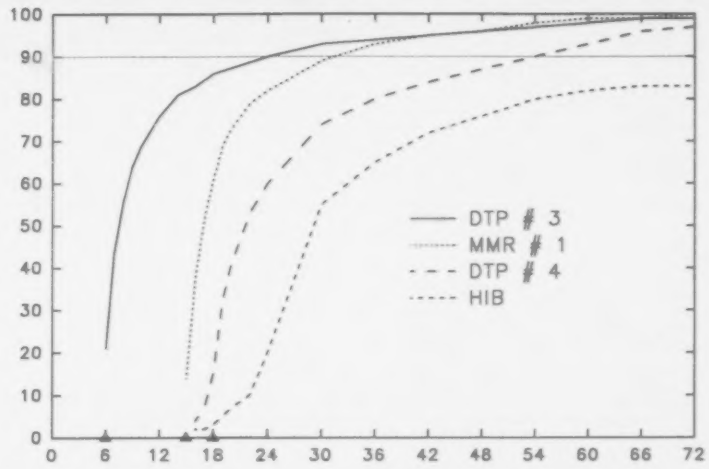
Requests for reprints should be sent to Lance E. Rodewald, MD, Box 655, Strong Memorial Hospital, Rochester, NY 14642-8655.

This paper was accepted May 11, 1993.

TABLE 1—Percentage of Study Children Receiving Vaccines by the Time of Their Entry Into School

Vaccine	Students Receiving Vaccine, %
DTP No. 1[a]	99.9
DTP No. 2[a]	99.7
DTP No. 3[a]	99.3
DTP No. 4[a]	96.1
DTP No. 5[b]	67.6
OPV No. 1[a]	99.8
OPV No. 2[a]	99.6
OPV No. 3[a]	98.8
OPV No. 4[b]	88.5
OPV No. 5[b]	44.1
MMR No. 1[a]	99.7
MMR No. 2[a]	90.5
Hib[b]	76.3

Note. DTP = diphtheria, tetanus, and pertussis; OPV = oral polio vaccine; MMR = measles, mumps, and rubella; Hib = *Hemophilus influenzae* type b.
[a]Required by New York State for school registration.
[b]Recommended by the American Academy of Pediatrics.



Note. DTP = diphtheria, tetanus, and pertussis; MMR = measles, mumps, and rubella; Hib = *Hemophilus influenzae* type b. The y axis shows the percentage of children who received the vaccine by the ages (in months) indicated on the x axis. The triangles indicate the recommended administration dates of the vaccines (6 months for DTP number 3, 15 months for MMR, and 18 months for DTP number 4 and Hib).

FIGURE 1—Pattern of receipt of selected vaccines.

Of the 2939 charts surveyed from the 1991 to 1992 class, 397 (13.5%) mentioned one of the two pediatric continuity clinics as the regular source of care. A sample of 171 charts, 97 from Strong Memorial Hospital and 74 from Rochester General Hospital, was drawn for cross checking. We tested for significant differences in vaccination rates between the samples drawn and the remainder (n = 226) and found none.

Because the survey did not use names, cross checks had to rely on matching by initials and date of birth. A chart not being found could be due to clerical reasons (name change, address change, or birth date inaccuracy) or to no chart being in the record room.

Of the 97 charts sought at Strong Memorial Hospital, 60 (62%) were located. Among these records, 766 immunizations were recorded in both the school survey and the medical record; 661 (86.3%) of these were identical in vaccine and date. The 105 (13.7%) vaccine discrepancies were as follows: 29 (3.8%) vaccinations were recorded in the school survey but not in the medical record, 40 (5.2%) were recorded in the medical record but not the school survey, and 36 (4.7%) had discrepancies in date of vaccination. Among the 36 discrepancies in dates, 14 were biased to make the school survey appear to have vaccines given at an older age and 22 were biased in the other direction.

Of the 74 charts sought at Rochester General Hospital, 54 (73%) were located. Among these records, 660 immunizations were recorded in both the school survey and the medical record; 562 (85.2%) of these were identical in vaccine and date. The 98 (14.8%) vaccine discrepancies were as follows: 37 (5.6%) vaccinations were recorded in the school survey but not in the medical record, 33 (5.0%) were recorded in the medical record but not the school survey, and 28 (4.2%) had discrepancies in date of vaccination. Among the 28 discrepancies in dates, 13 were biased to make the school survey appear to have vaccines given at an older age and 15 were biased in the other direction.

We then checked for differences in vaccination status between the school records found and those not found at the two continuity clinics. At the Rochester General Hospital clinic, no statistically significant differences were found. At the Strong Memorial Hospital clinic, however, where fewer charts could be located, we found that children with locatable charts had higher vaccination rates than did children for whom no chart could be found. We could not determine whether the primary care provider was not listed correctly for these children or whether one of the variables used to match the school record with the medical chart was incorrect.

Table 1 shows the number of specific vaccinations per child for both survey years, and Figure 1 shows the pattern of receipt of selected individual vaccines. It can be seen that 18 months from the recommended administration date[5] were required to achieve a 90% vaccination rate for DTP number 3. Similarly, 90% vaccination rates required 36 months for DTP number 4 and 15 months for MMR. A 90% vaccination rate was not attained for the Hib vaccine. The mean length of time necessary to achieve 90% vaccination rates for DTP numbers 1 through 4 and MMR was 14.2 months.

Discussion

School-based immunization surveys gather patient-level details about vaccination histories during the preschool years of life, the time when children are most vulnerable to vaccine-preventable diseases.[6,7] We have shown that these vaccination histories can be reconstructed in a simple computer database to facilitate the calculation of several measures of vaccination status.[8]

Error in school-based immunization survey data can be introduced at several levels, most perniciously by inaccurate or incomplete recording of dates by school officials. Such error would lower calculated rates. Our data suggest that relative to medical chart review, error occurred in both directions and was therefore unbiased.

This survey methodology is not only powerful but inexpensive. The entire sur-

vey was conducted in a 3-week period by three chart abstractors. Once permission is obtained to survey the school district, the actual implementation of the survey is not difficult. The algorithms for measuring vaccination rates are straightforward and can be encoded in either database or statistical package languages.

Because school-based data include children who have no identified provider or who change providers, these surveys are useful for identifying groups of children who may lack connections to the primary care system. Because these children are at the highest risk of adverse outcome, inadequate screening for disease, and inadequate vaccination, this methodology can potentially identify targets for interventions. In contrast, provider-based surveys cannot identify children with inadequate access to the primary care system. A characteristic of school-based immunization surveys, however, is that the surveyed population consists of those children who remain in the city and attend a city school. Migration from the city will reduce the survey scope as a result of omission of those children who were city residents during their preschool years but who moved to the suburbs by school entry. Busing children into or out of the school district will also complicate the survey.

Closing the "toddler gap" in vaccination services will probably require a multifaceted approach. One of the essential early steps is to obtain data that can help to target interventions accurately. Retrospective school-based immunization surveys can provide those data in a short period of time and with minimum effort. □

Acknowledgments

This work was supported by a grant from the Centers for Disease Control (200-90-0869).

This research was presented at the 26th National Immunization Conference, June 1992, St. Louis, Mo.

We thank Andrew Doniger, MD, and the Rochester City School District for their support. We thank the project staff for their efforts.

References

1. Centers for Disease Control. Measles vaccination levels among selected groups of preschool-aged children—United States. *MMWR.* 1991;40:36–39.
2. Centers for Disease Control. Retrospective assessment of vaccination coverage among school-aged children—selected U.S. cities, 1991. *MMWR.* 1992;41:103–107.
3. Orenstein WA, Bernier RH. Surveillance: information for action. *Pediatr Clin North Am.* 1990;37:709–734.
4. *Sampling Procedure for Conducting Immunization Assessment/Validation Surveys for School and Day-Care Centers—Retrospective Surveys Using School Systems Databases and Guidelines for Public Health Immunization Clinic Audits for Immunization Project Areas.* Atlanta, Ga: Centers for Disease Control; 1990.
5. Committee on Infectious Diseases. *1988 Red Book.* Elk Grove Village, Ill: American Academy of Pediatrics; 1988.
6. Guyer B, Baird SJ, Hutcheson RH, et al. Failure to vaccinate children against measles during the second year of life. *Public Health Rep.* 1976;91:133–137.
7. Cutts FT. Failure to vaccinate. In: *Proceedings of the 24th National Immunization Conference.* Atlanta, Ga: Centers for Disease Control; 1990:55–58.
8. Raubertas RF, Rodewald LE, Humiston SG, Szilagyi PG. Defining and measuring undervaccination. Presented at the 25th National Immunization Conference, June 1990, Washington, DC.

The Impact of a Public Health Nurse Intervention on Influenza Vaccine Acceptance

Margaret E. Black, MSN, Jenny Ploeg, MScN, Stephen D. Walter, PhD, ARCS, Brian G. Hutchison, MD, MSc, CCFP, E. A. Fran Scott, MD, MSc, CCFP, FRCP, and Larry W. Chambers, PhD, MSc

ABSTRACT

Public health clients in an Ontario community 65 years of age or older were randomly allocated to receive an intervention by a public health nurse during a home visit promoting either influenza immunization or safety measures. There was no statistically significant difference in influenza immunization rates between these two groups (56.1% vs 56.6%). Men were significantly more likely to receive immunization. (*Am J Public Health.* 1993;83:1751–1753)

Introduction

Annual vaccination against influenza is currently recommended for all persons 65 years of age or older.[1–4] Despite this recommendation, vaccination rates among community-living North American seniors remain at approximately 20% to 30%.[5–7] The literature suggests that structured personalized communication[8–14] using components of the health belief model and expectancy-value theory offers considerable promise in altering health behaviors. The current lack of incentives for primary care physicians to promote vaccination and the positive experience in nurse-led interventions[15–17] provide encouragement for the development of a public health nurse intervention.

The authors are with the Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada. Margaret E. Black, Jenny Ploeg, Stephen D. Walter, E. A. Fran Scott, and Larry W. Chambers are also with the Regional Municipality of Hamilton-Wentworth Department of Public Health Services, Hamilton. Brian G. Hutchison is also with the First Place Family Medical Centre, Hamilton.

Requests for reprints should be sent to Jenny Ploeg, MScN, Hamilton-Wentworth Department of Public Health Services, 25 Main St W, Suite 400, PO Box 897, Hamilton, Ontario, Canada L8N 3P6.

This paper was accepted August 30, 1993.

The results and conclusions presented here are the authors'. No official endorsement by the Ontario Ministry of Health is intended or inferred.

TABLE 1—Comparison of Promotion Groups (n = 359)

	Influenza (n = 204)	Safety (n = 155)
Sex, no. (%)		
Men (n = 117)	59 (28.9)	58 (37.4)
Women (n = 242)	145 (71.1)	97 (62.6)
Mean age, y (SD)	77.59 (8.4)	77.78 (6.5)
At least 1 chronic health problem reported, no. (%)	134 (65.7)	104 (67.1)
Alternative and assigned promotion received, no. (%)	15 (8.1)	9 (6.1)
Self-rated health, mean* (SD)	2.55 (0.89)	2.70 (0.87)

*1 = excellent, 4 = poor.

TABLE 2—Frequency of Responses to Influenza Immunization Items

	Influenza (n = 198), %	Safety (n = 152), %	Difference between Groups, % (95% Confidence Interval)
Reported receipt of immunization	56.1	56.6	−0.5 (−11.0, 10.0)
Reported talking about immunization with public health nurse	42.2	18.2	24.0 (13.9, 31.9)

Methods

Research Design and Setting

Public health clients were randomly assigned to receive a public health nurse intervention during a home visit promoting either influenza immunization (study group) or safety measures (control group). The study took place in an urban–rural region of 450 000 residents, 13.3% of whom were 65 years of age or older.[18]

All 1011 clients aged 65 years or older who were considered active clients as of August 1990 or were referred to public health between September 1, 1990, and January 15, 1992, were randomly assigned to one of two groups. They were visited or contacted by 1 of the 18 seniors program public health nurses to determine, during the course of their usual care, whether they were eligible and willing to participate in the study. If it was determined that the client and/or proxy (family or close friend) were not able to or could or would not participate, the public health nurse did not proceed with the promotion and the client was therefore not part of the study. Some clients required more than one visit for the promotion. The public health nurses were instructed to provide their "usual care" regarding the topic of the alternate promotion.

The influenza promotion was designed to be individualized to address each client's unique concerns. The public health nurse reviewed influenza and its vaccine with clients and identified strategies to overcome barriers to immunization by family physicians (e.g., transportation services available in the community).

Data Collection

Outcome data were obtained through telephone interview (or home visit) by two research assistants who were unaware of group membership. Assignment of clients to research assistants was randomized. All clients were asked identical questions regarding the receipt of influenza immunization, self-rated health, and the presence of high-risk conditions[19,20] (heart disease, respiratory disease, or diabetes). Clients who did not receive the immunization were asked to explain their reasons. Both client groups were asked safety questions.

Results

Of the 1011 potential study clients, 589 (58.0%) were excluded for reasons such as cognitive impairment and "inactive" client status. An additional 57 clients (5.6%) refused the promotion, and 6 (0.6%) were lost to follow-up. Of those who were recruited, 157 received the influenza promotion and 148 received the safety promotion. Another 45 clients had been assigned to the influenza group but did not receive the promotion because the public health nurse found that they had already been administered influenza vaccine. These 45 subjects and those who were missed (n = 9) were included in the analysis in their originally allocated group (an "intention to treat"[21] analysis); thus, a total sample of 359 was analyzed.

Among these 359 clients, 117 were men and 242 were women; the mean age was 77.2 years. The sex and age distributions were very similar in the two groups (see Table 1). No between-group differences were found in the proportion of subjects reporting at least one chronic health condition or receiving the alternate promotion together with the assigned promotion. The mean ratings of self-rated health were similar.

At follow-up, there were no differences between groups in the number of subjects interviewed through a proxy, the number assigned to the two interviewers, and the number deceased. Also, there were no statistically significant differences in the rates of self-reported influenza immunization in the influenza and safety groups (56.1% and 56.6%, respectively) (see Table 2).

A post hoc power analysis confirmed that the sample size was sufficient to have detected a 50% relative increase in the vaccination rate (80% power, 5% alpha, two-sided). After the effects of the presence of chronic health conditions, self-rated health, and sex were controlled, there was still no statistically significant difference between the group immunization rates. In a pooled analysis of both study groups, we found that men were more likely than women to receive influenza immunization (70.5% and 49.6%, respectively).

Multiple logistic regression was used to examine the relative contribution of sex, promotion group, age, chronic health conditions, self-rated health, public health nurse, and public health nurse experience on the likelihood of receiving the influenza immunization. Even when group membership was forced into the model, only the effect of sex was significantly associated with the likelihood of receiving immunization when all of the variables were considered simultaneously ($\beta = .89$ [SE] .25), $P < .003$, 95% CI = 1.51, 4.01).

Forty-nine clients (33.7% and 45.7% of the nonvaccinated clients in the influenza and safety groups, respectively) gave reasons for not receiving immunization. The three most frequent reasons in the influenza promotion group were forgetting/confusion, prior experience with vaccine side effects, and beliefs regarding immuni-

zation. In the safety group, the most frequent reasons were beliefs regarding immunization and previous experience with vaccine side effects.

Discussion

The immunization rates reported here were considerably higher than the expected community rates of 20% to 30% for older persons.[5–7] Because of the nature of public health referrals, our subjects may be more likely to see a physician regularly and may be perceived as at greater risk; therefore, they may be more likely to be offered immunization. However, because there are no recent surveys on immunization rates among older adults in our community, we cannot be certain of the extent to which this subgroup is representative of the older community in general.

The reasons clients reported for failing to receive immunization were consistent with those previously reported[14,22] and appear to be an important barrier to preventive behavior. In contrast to findings reported by Conn,[23] the presence of chronic health conditions and self-rated health were not related to immunization behavior in this study.

Our finding that men were more likely than women to receive immunization was puzzling. Future studies could investigate population- and sex-specific influenza immunization rates. Researchers could also examine the effect of peer support in overcoming the reluctance of those whose beliefs constitute barriers to vaccine acceptance. Finally, qualitative research methods may contribute to our understanding of older clients' perceptions and their willingness to participate in preventive self-care. □

Acknowledgments

This project was funded by the Ontario Ministry of Health, Health Care Systems Research Program.

We express our appreciation to Carolynne Turner for her technical assistance, including data entry; to Christine Walsh and Lois Alldis for conducting the interviews; to Derek Watter for data analysis; and to the Senior Program public health nurses for their cooperation and commitment to this study.

References

1. Immunization Practices Advisory Committee. Prevention and control of influenza. *MMWR.* 1991;40:1–15.
2. National Advisory Committee on Immunization. Statement on influenza vaccination for the 1991–1992 season. *Can Dis Weekly Rep.* June 15, 1991:17–24, 121–126.
3. LaForce FM. US Preventive Services Task Force: immunizations, immunoprophylaxis, and chemoprophylaxes to prevent selected infections. *JAMA.* 1987;257: 2464–2470.
4. Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J.* 1979;121: 1193–1254.
5. Kavet J. Vaccine utilization: trends in the implementation of public policy in the USA. In: Selby P, ed. *Influenza: Virus, Vaccines and Strategy.* New York, NY: Academic Press; 1976:297–308.
6. Fedson DS. Influenza and pneumococcal immunization strategies for physicians. *Chest.* 1987;91:436–443.
7. Influenza vaccination levels in selected states—Behavioral Risk Factor Surveillance System, 1987. *MMWR.* 1989;38:124–133.
8. Larson EB, Olsen E, Cole W, et al. The relationship of health beliefs and a postcard reminder to influenza vaccination. *J Fam Pract.* 1979;8:1207–1211.
9. Frank JW, Henderson M, McMurray L. Influenza vaccinations in the elderly: 1. Determinants of acceptance. *Can Med Assoc J.* 1985;132:371–375.
10. Cummings KM, Jette AM, Brock BM, et al. Psychosocial determinants of immunization behavior in a swine influenza campaign. *Med Care.* 1979;17:639–649.
11. Riddiough MA, Willems JS, Sanders CR, et al. Factors affecting the use of vaccines: considerations for immunization program planners. *Public Health Rep.* 1981;96:528–535.
12. Montano DE. Predicting and understanding influenza vaccination behavior: alternatives to the health belief model. *Med Care.* 1986;24:438–453.
13. Carter WB, Beach LR, Inui TS, et al. Developing and testing a decision model for predicting influenza vaccination compliance. *Health Serv Res.* 1986;20:897–932.
14. Buffington J, Bell KM, LaForce FM, et al. A target-based model for increasing influenza immunizations in private practice. *J Gen Int Med.* 1991;6:204–209.
15. Hoey JR, McCallum HP, LePage EMM. Expanding the nurse's role to improve preventive service in an outpatient clinic. *Can Med Assoc J.* 1982;127:27–28.
16. Fedson DS, Kessler HA. A hospital-based influenza immunization program 1977–78. *Am J Public Health.* 1983;73:442–445.
17. Ratner ER, Fedson DS. Influenza and pneumococcal immunization in medical clinics, 1978–1980. *Arch Intern Med.* 1983; 143:2066–2069.
18. *Hamilton-Wentworth Population Projections 1988–2006.* Hamilton, Ontario, Canada: Hamilton-Wentworth Planning and Development Department, Strategic Planning Division; 1989.
19. Naus M. Summary report of the Ontario influenza season, 1989–1990. *Ontario Dis Surveill Rep.* 1990;11:109–120.
20. National Advisory Committee on Immunization. Statement on influenza vaccination for the 1990–91 season. *Can Dis Weekly Rep.* 1990;16:117–122.
21. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol.* 1992;21:837–841.
22. Pocock SJ. *Clinical Trials.* New York, NY: Wiley; 1988:76–79.
23. Conn V. Self-care actions taken by older adults for influenza and colds. *Nurs Res.* 1991;40:176–181.

Neonatal Tetanus in Peru: Risk Assessment with Modified Enzyme-Linked Immunosorbent Assay and Toxoid Skin Test

Louis Vernacchio, BA, Guillermo Madico, MD, Manuela Verastegui, MS, Fernando Diaz, MD, Timothy S. Collins, BA, and Robert H. Gilman, MD

ABSTRACT

We used a modified enzyme-linked immunosorbent assay (ELISA) to investigate tetanus immunity in 232 pregnant Peruvian women. One hundred forty-two (61.2%) had protective antitoxin titers (≥0.01 IU/mL). Protective titers correlated positively with the number of toxoid doses reported during the current pregnancy. A majority of women reporting no toxoid doses during the current pregnancy had at least one prenatal health care visit. We evaluated a toxoid skin test in 44 of the subjects, but it correlated poorly with the ELISA. The modified ELISA is a useful in vitro method for studying tetanus immunity in the developing world. (*Am J Public Health.* 1993;83:1754–1756)

Introduction

Neonatal tetanus causes over 780 000 deaths worldwide each year.[1] Peru reports approximately 0.4 cases per 1000 live births, one of the highest rates in Latin America.[2] Yet adequate maternal immunization with tetanus toxoid can completely prevent the disease.[3,4] During pregnancy, two adequately spaced injections of tetanus toxoid induce protective levels of antitoxin in more than 80% of previously unimmunized women, and three injections protect virtually 100%.[5,6]

Antitetanus vaccination efforts in the Third World have failed to reach large numbers of women of childbearing age. No information on the success of such efforts in Peru has been published.

Serologic investigation of tetanus immunity in the developing world presents several problems. The gold standard mouse neutralization assay is costly and cumbersome. Routine use of the enzyme-linked immunosorbent assay (ELISA) and hemagglutination have proven unreliable for measuring low titers.[7,8] Simonsen et al. have developed a modified competitive ELISA that correlates extremely well with the in vivo standard.[9] This recently developed in vitro assay has not been used to study tetanus immunity in a Third World population.

Even with the modified ELISA, large-scale studies of tetanus immunity in the developing world remain difficult. Intradermal injections of tetanus toxoid elicit a delayed hypersensitivity reaction in a large percentage of previously immunized subjects, and this reaction may correlate with serum antitoxin titer.[10,11] A reliable skin test would greatly facilitate the study of tetanus immunity.

We designed a study to assess serum tetanus antitoxin levels in pregnant Peruvian women with the modified ELISA. We administered a tetanus toxoid skin test to a subset of our subjects to investigate its use as a screen for tetanus immunity.

Subjects and Methods

Subjects

Subjects were pregnant women presenting in labor to three hospitals: Cayetano Heredia, serving a poor and lower-middle-class population in Lima; Maternidad de Lima, a large maternity hospital serving poor women from Lima and its shantytowns; and Hospital Regional del Amazonas in the city of Iquitos in the Amazon jungle.

Women admitted in labor to these hospitals during the study period were enrolled. Verbal consent was obtained. Subjects were interviewed about antitetanus vaccination during their current pregnancy. A blood specimen was obtained by venipuncture. A tetanus toxoid skin test was performed on a subset of the subjects.

ELISA

The modified ELISA was performed according to Simonsen et al.[9]

Skin Test

A 0.1-mL dose of vaccine-grade adsorbed tetanus toxoid (Connaught) was injected intradermally on the anterior aspect of the forearm. Reactions were scored as the greatest diameter of induration at 36 to 72 hours. A positive cutoff of 5 mm or more was chosen.

Louis Vernacchio and Timothy S. Collins are with the Johns Hopkins University School of Medicine, Baltimore, Md. Guillermo Madico, Manuela Verastegui, and Fernando Diaz are with the Universidad Cayetano Heredia and A.B. PRISMA, Lima, Peru. Robert H. Gilman is with the Department of International Health, The Johns Hopkins University School of Hygiene and Public Health.

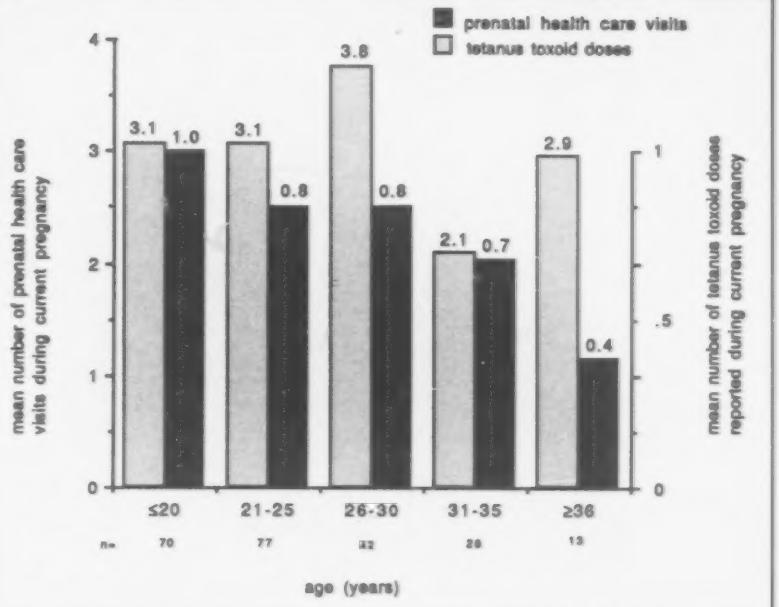
Requests for reprints should be sent to Robert H. Gilman, MD, Department of International Health, The Johns Hopkins University School of Hygiene and Public Health, 615 N Wolfe St, Baltimore, MD 21205.

This paper was accepted August 10, 1993.

TABLE 1—Percentage of Subjects with Protective Serum Tetanus Antitoxin Levels, by Hospital

Hospital (Location)	No. Subjects	Subjects with Protective Antitoxin Level (≥ 0.01 IU/mL),[a] %
Maternidad (capital)	141	58.2
Cayetano (capital)	21	71.4
Amazonas (jungle)	70	64.3
Total	232	61.2

[a] $P = 0.42$, chi-square test, for comparison among hospitals.



Note. n = 230 because prenatal health care data were not available for two subjects. Prenatal healthcare visits, $P = .69$, linear regression analysis; tetanus toxoid doses, $P = .027$, linear regression analysis.

FIGURE 1—Mean number of prenatal health care visits and mean number of tetanus toxoid doses reported during the current pregnancy, by age group.

Results

A total of 232 pregnant women in labor (mean age ± SD = 24.5 ± 5.9 years) were included in the study. The number of pregnancies (including the current one) ranged from 1 to 10 (median = 2; 10th and 90th percentiles = 1 and 6). Subjects reported a mean of 3.1 (SD = 3.0) prenatal health visits and a mean of 0.9 (SD = 0.9, range = 0–3) tetanus toxoid injections during the current pregnancy.

A total of 142 of 232 (61.2%) subjects had a serum tetanus antitoxin concentration above the protective level of 0.01 IU/mL. No difference was found among hospitals ($P = .42$, chi-square test; Table 1).

Of the subjects reporting 0, 1, 2, and 3 tetanus toxoid injections during the index pregnancy, 39.6%, 71.7%, 87.3%, and 100%, respectively, had a protective serum antitoxin level ($P < .00001$, chi-square test for linear trend). Both the percentage of subjects with positive titers and the number of tetanus toxoid doses reported during the pregnancy decreased with increasing age group ($P = .00036$, chi-square test for linear trend [positive titers]; $P = .027$, linear regression analysis [tetanus toxoid doses]), but we found no relationship between age and the number of prenatal health care visits ($P = .69$, linear regression analysis; Figure 1).

Of the subjects who reported receiving no tetanus toxoid injections during the pregnancy, 55 of 106 (51.9%) gave a history of at least one prenatal health care visit and 49 of 106 (46.2%) gave a history of two or more such visits. Of those reporting only one injection, 39 of 60 (65.0%) reported at least two prenatal health care visits.

Tetanus toxoid skin tests were administered to 70 subjects. Twenty-six tests (37.1%) were not read because the subject was discharged less than 36 hours after placement. Of the 44 skin tests read, 18 (40.9%) were positive, but the test correlated poorly with the modified ELISA, with a sensitivity of 57.7% and a specificity of 83.3% (Figure 2).

Discussion

The modified ELISA developed by Simonsen et al.[9] was useful for studying tetanus immunity in a developing country and provides an alternative to the standard in vivo assay.

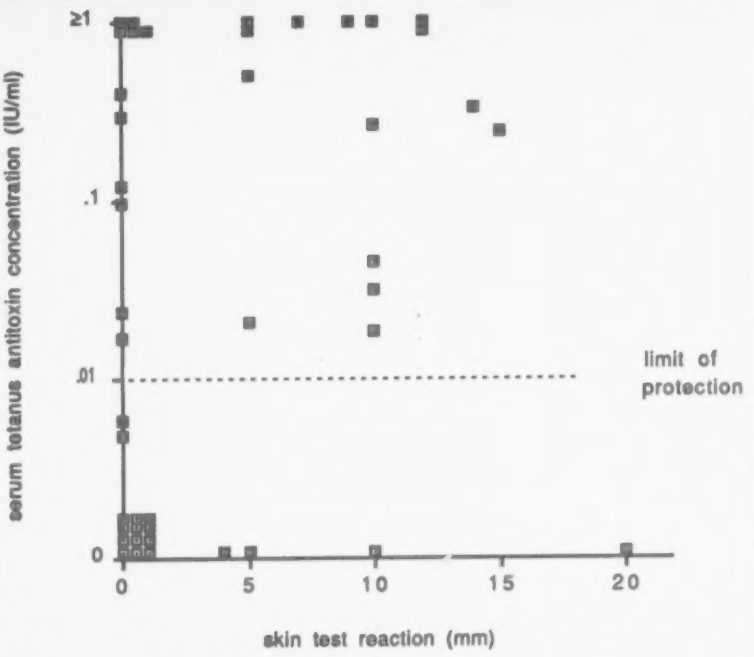
Roughly 40% of hospitalized pregnant Peruvian women lack sufficient serum tetanus antitoxin to protect their newborns against neonatal tetanus. That the rate in a jungle hospital was comparable to that of two Lima hospitals probably reflects recent vaccination efforts to target the jungle region of Peru, an area traditionally endemic for tetanus.

The overall 60% rate of protection exceeds the 1990 World Health Organization estimate of 29% for pregnant women in the developing world.[1] Our survey is, however, limited to hospitalized women. We would expect the coverage among women giving birth outside of hospitals to be lower owing to decreased access to health care.

A striking finding of our study was that a large percentage of women receiving no tetanus toxoid injections during their pregnancy reported one and often more prenatal health care visits. Clearly, many opportunities for vaccination were missed in our study population.

Our results confirm the highly significant correlation between the number of tetanus toxoid doses during pregnancy and the protective antitoxin titers. Among those subjects reporting two or three vaccinations during the pregnancy, almost 90% had protective antibody titers. Yet 40% of the subjects reporting no vaccinations during the pregnancy nevertheless had protective titers. Their immunity may represent either previous vaccination or naturally acquired antibody. Data from India and Brazil indicate that 3% to 40% of nonvaccinated subjects have positive antitoxin titers.[12–14] These findings suggest that health care workers can accept a patient's report of recent vaccinations, although a report of no recent vaccinations does not rule out immunity.

Younger women in our study were vaccinated more frequently than older women. Because we found no relation-



Note. Minimum positive skin test reaction was ≥5 mm: sensitivity, 57.7%; specificity, 83.3%. $R = .161$; $P = .30$, linear regression analysis.

FIGURE 2—Serum tetanus antitoxin level as determined by the modified ELISA vs the maximum diameter of induration with the tetanus toxoid skin test.

ship between age and the number of prenatal health care visits, the younger women either were offered vaccination more frequently or accepted it more often.

The tetanus toxoid skin test was of little value as a predictor of tetanus immunity. Factors responsible for the many false-negative results may include a decrease in cellular immune response to specific antigens in pregnant women near term[15,16] and the time period between vaccination and skin testing.

Neonatal tetanus remains an important cause of neonatal mortality in the developing world. Even in a population of Peruvian women with access to health care, tetanus vaccination rates remain far from ideal. Further studies to determine the reasons for missed vaccination opportunities would be helpful.

The modified ELISA was useful for investigating tetanus immunity in a developing country. Its wider application would facilitate study of other Third World populations. □

Acknowledgments

Louis Vernacchio was granted the Harold Lamport Student Biomedical Research Award from The Johns Hopkins University School of Medicine for this research.

The authors thank the directors and staffs of Hospital Cayetano Heredia, Hospital Maternidad de Lima, and Hospital Regional del Amazonas; the Lederle company for their donation of tetanus toxoid; and Dr Robert Black for review of the manuscript.

References

1. Global Advisory Group. Expanded program on immunization: part I. *Weekly Epidemiol Rev.* 1990;65:1–11.
2. da Silveira CM, de Quadros CA. Neonatal tetanus: countdown to 1995. *World Health Forum.* 1991;12:289–296.
3. Newell KW, Dueñas Lehmann A, Leblanc DR, Garces Osorio N. The use of toxoid for the prevention of tetanus neonatorum. *Bull World Health Organ.* 1966; 35:863–871.
4. Black RE, Huber DH, Curlin GT. Reduction of neonatal tetanus by mass immunization of non-pregnant women: duration of protection provided by one or two doses of aluminium-adsorbed tetanus toxoid. *Bull World Health Organ.* 1980;58:927–930.
5. Chen ST, Edsall G, Peel MM, Sinnathuray TA. Timing of antenatal tetanus immunization for effective protection of the neonate. *Bull World Health Organ.* 1983;61: 159–165.
6. Rethy L, Rethy LA. Active anti-tetanus immunisation of females to control neonatal tetanus. *Lancet.* 1986;1:616. Letter.
7. Simonsen O, Bentzon MW, Heron I. ELISA for the routine determination of antitoxic immunity to tetanus. *J Biol Standardization.* 1986;14:231–239.
8. Simonsen O, Kjeldsen K, Heron I. Immunity against tetanus and effect of revaccination 25–30 years after primary vaccination. *Lancet.* 1984;2:1240–1242.
9. Simonsen O, Schou C, Heron I. Modification of the ELISA for the estimation of tetanus antitoxin in human sera. *J Biol Standardization.* 1987;15:143–157.
10. Facktor MA, Bernstein RA, Fireman P. Hypersensitivity to tetanus toxoid. *J Allergy Clin Immunol.* 1973;52:1–12.
11. Bhatia S, Ray K, Mathur R, et al. Evaluation of cutaneous response to tetanus antigen to document immunization: preliminary observations. *Indian Pediatr.* 1985;22:267–273.
12. Dastur FD, Awatramani VP, Dixit SK, D'Sa JA, Cooverji ND, Anand MP. Response to single dose of tetanus vaccine in subjects with naturally acquired tetanus immunity. *Lancet.* 1981;2:219–221.
13. Ray SN, Ray K, Grover SS, Sharma RS, Sharma SP. Sero-survey of diphtheria and tetanus antitoxin. *Indian J Med Res.* 1978; 68:901–904.
14. Veronesi R, Cecin H, Correa A, Tavares J, Moraes C, Bertoldo OJ. New concepts on tetanus immunization: naturally acquired immunity. *J Hyg Epidemiol Microbiol Immunol.* 1975;19:126–134.
15. Sridama V, Pacini F, Yang S, Moawad A, Reilly M, DeGroot LJ. Decreased levels of helper T cells in pregnancy: a possible cause of immunodeficiency in pregnancy. *N Engl J Med.* 1982;307:352–356.
16. Thong YH, Steele RW, Vincent MM, Hensen SA, Bellanti JA. Impaired in vitro cell-mediated immunity to rubella virus during pregnancy. *N Engl J Med.* 1973;289:604–606.

HIV Transmission: Women's Risk from Bisexual Men

Robert W. Wood, MD, Leigh E. Krueger, MPH, Tsilke C. Pearlman, and Gary Goldbaum, MD, MPH

ABSTRACT

The purpose of this study was to characterize the risk of human immunodeficiency virus (HIV) infection for men who have sex with men and to identify the risk such men pose to their female sex partners. The subjects were 5480 men who were tested for HIV between January 1987 and December 1991 and who reported having had sex with a man since 1977. Men who identified themselves as bisexual or straight were more likely to use injection drugs, had a substantial HIV seroprevalence, and reported many more female partners than men who identified themselves as gay. Men who identify themselves as bisexual pose the greatest risk to their female partners. (*Am J Public Health.* 1993;83:1757–1759)

Introduction

A recent analysis of national acquired immunodeficiency syndrome (AIDS) surveillance data indicated that men who reported having sex with both men and women (bisexuals) differed from those who reported having sex only with men (homosexuals) in important ways.[1] Although two thirds (67%) of women with heterosexually acquired AIDS (excluding women born in the Caribbean or Africa) reported that their sex partner was an injection drug user,[2] 11% reported sex with a bisexual man as their only risk.[1] Because bisexual men are increasingly likely to contribute to human immunodeficiency virus (HIV) infections in women,[1,3–6] we sought to further characterize the risk of HIV infection for men who have sex with men and to identify the risk such men pose to women.

Methods

The AIDS Prevention Project of the Seattle–King County Department of Public Health actively recruits for HIV counseling and testing persons at highest risk for HIV. Although the Project accounts for only 23% of the HIV testing performed at the Seattle–King County Department of Public Health, it identifies more seropositive persons than all other Department clinics combined (203 of 352 [58%] in 1991).

Trained staff collect standard demographic, medical, and risk data on all persons seeking HIV counseling and testing, including the number of male and female sex partners during the previous 12 months, self-stated sexual identity ("gay," "bisexual," "straight"), specific sexual practices and condom use, and other risks for HIV infection, including injection drug use. For HIV testing, blood undergoes enzyme immunoassay analysis (LAV EIA, Genetic Systems, Seattle, Wash),[7] which is repeated if reactive and confirmed by indirect fluorescence assay (VIRGO HIV IFA, Pharmacia Diagnostics, Inc, Columbia, Md)[8] or Western blot (Biotech/DuPont HIV-1 Western Blot Kit, Ortho Diagnostic Systems Inc, Raritan, NJ) analysis.[9]

We analyzed cross-sectional data from each subject's latest clinic visit between January 1987 and December 1991. The subjects were 5480 men who reported having sex with another man since 1977 and who elected to undergo HIV testing following pretest counseling. (Only 3.2% of men counseled elected not to test.) We analyzed data separately for HIV-infected men, whether or not they reported having sex during the previous year, and uninfected men who reported having sex with both men and women in the past year.

For categorical data we calculated the chi-square or Fisher's Exact Test, plus odds ratios with 95% confidence intervals.[10] Differences are considered significant if $P < .05$.

Results

Self-reported sexual identity was an important predictor of having sex with women and of HIV seropositivity at our site. Of the 5480 subjects, 77% identified themselves as gay and reported almost no sex with women in the previous year (average 0.07 female partners); 13% identified themselves as bisexual (average 1.4 female partners); and 8.4% identified themselves as straight (average 2.7 female partners). Men who self-identified as gay had the highest HIV seroprevalence (27%); however, substantial numbers of those who said they were bisexual or straight were also infected (12% and 8%, respectively).

Injection drug use was reported by 11% of the gay men, 20% of the bisexual men, and 34% of the straight men ($P < .05$). Injection drug users reported

Robert W. Wood, Leigh E. Krueger, and Tsilke C. Pearlman are with the AIDS Prevention Project, Seattle–King County Department of Public Health, Seattle, Wash. Robert W. Wood is also with the School of Medicine, University of Washington, Seattle. Gary Goldbaum is with the School of Public Health and Community Medicine, University of Washington, Seattle.

Requests for reprints should be sent to Robert W. Wood, MD, AIDS Prevention Project, 2124 4th Ave, 4th Floor, Seattle, WA 98121.

This paper was accepted March 12, 1993.

TABLE 1—HIV Seroprevalence by Sexual Identity, and Exposure of Female Partners by HIV-Infected Men Who Have Sex With Men: Seattle, Wash, 1987 through 1991

	n	HIV Prevalence, %	HIV-Infected Men with Female Partners		
			No. Men	%[a]	Total No. Female Partners
"Straight" men	457	7.7	12	34.3	32
Non-IDUs	303	7.6	4	17.4	5
IDUs	154	7.8	8	66.7	27
"Bisexual" men	711	12.4	31	35.2	116
Non-IDUs	567	9.5	15	27.8	38
IDUs	144	23.6	16	47.1	78
"Gay" men	4226	26.6	18	1.6	32
Non-IDUs	3772	24.5	12	1.3	22
IDUs	454	44.5	6	3.0	10
Total	5480	23.0	63	5.0	182
Non-IDUs	4712	21.4	32	3.2	66
IDUs	768	33.1	31	12.2	116

Note. IDU = injection drug user. The table totals include data on 86 persons (16 IDUs) who did not specify their sexual identity.
[a]Percentages are based on seropositive men.

TABLE 2—Number of Sex Partners of Uninfected Bisexually Active Men in the Preceding 12 Months, by Stated Sexual Identity and Injection Drug Use

	n	Mean No. Partners per Subject	
		Female	Male
"Straight" men	70	3.8	2.8
Non-IDUs	52	3.0	2.0
IDUs	18	6.1	5.3
"Bisexual" men	277	2.6	7.9
Non-IDUs	222	2.0	6.5
IDUs	55	4.8	13.6
"Gay" men	86	2.3	9.9
Non-IDUs	76	2.3	10.0
IDUs	10	2.2	9.4
Total	451	2.8	7.7
Non-IDUs	363	2.2	6.8
IDUs	88	4.9	11.5

Note. IDU = injection drug user.

having more female sex partners during the previous 12 months (mean 1.5) than did those who did not report injection drug use (mean 0.3). Injection drug users were also more likely than nonusers to be HIV-infected (33% vs 21%) (odds ratio [OR] = 1.8, 95% confidence interval [CI] = 1.5, 2.2).

Of the seropositive men, most (89%) identified themselves as gay, but most (81%) of the female partners exposed in the previous year were exposed by straight and bisexual men (Table 1). Significantly more injection drug users than nonusers reported having female partners (OR = 4.2, 95% CI = 2.4, 7.4). Eleven (17%) of 63 HIV seropositive men reported always using condoms for vaginal sex during the previous 6 months; 15 (24%) reported having only unprotected vaginal sex. Injection drug users were more likely than nonusers to report having unprotected vaginal sex, a nonsignificant difference.

Of the 451 HIV-seronegative men who reported having had both male and female partners in the previous year, almost two thirds (277, or 61%) identified themselves as bisexual (Table 2). Although only one in five (20%) of these men were injection drug users, those who were injection drug users averaged more than twice as many female sex partners as the seronegative bisexual men who did not use injection drugs. These bisexual men reported always using condoms only 12% of the time, and 30% reported never using condoms. Injection drug users did not differ from nonusers in their use of condoms.

Discussion

Compared with men who identified themselves as gay or bisexual, straight-identified men who have sex with men reported having more female sex partners and fewer male partners during the preceding year, and they were less likely to be HIV-infected. Even so, HIV seroprevalence among straight-identified men who have sex with men was remarkably high (8%), although lower than in bisexual men (12%). Straight men reported almost twice as many female sexual partners as bisexual men and 38 times as many female partners as gay men. Men who identified themselves as bisexual exposed the most women to HIV infection in the previous 12 months; gay and straight men exposed equal numbers of women. Women's risk is greatest from self-identified bisexual men. Men who have sex with men, both HIV seropositive and seronegative, had a low rate of consistent condom use during vaginal sex (12%).

Our conclusions pertain to Seattle men who have sex with men seeking HIV counseling and testing at an AIDS prevention program and may not reflect the population of men who have sex with men at large. Although a fifth of our subjects identified themselves as straight or as bisexual, this proportion could be too low if straight-identified men who have sex with men are reluctant to come to a gay-oriented AIDS clinic or do not perceive themselves to be at risk for HIV infection. Recruiting such men may require different strategies.[11]

Sexual self-identification may be a geographically variable phenomenon; our findings should be studied in other locales. Reporting biases might result if gay men underreport female partners or bisexual or straight men overreport them. Interviewer biases (since some are gay and more attuned to gay men) could result in incomplete ascertainment of female partners (and their risk). The use of standard protocols for interviews and data recording, however, should reduce this source of bias.

As others have noted,[1] the population of men who have sex with men but do not identify themselves as gay is in great need of further study to determine its size, its risks for further HIV transmission, and the best intervention methods. In the meantime, it appears that increased attention should focus on men who have sex with men and who also use injection drugs. □

Acknowledgments

This study was supported by the AIDS Prevention Project, Seattle–King County Department of Public Health.

The authors would like to express thanks to Bill Ford, who provided secretarial and editing assistance; Larry Keil, who helped locate citations; and Paul Swenson, PhD, whose lab performed the HIV testing and who documented the testing process.

References

1. Chu SY, Peterman TA, Doll LS, et al. AIDS in bisexual men in the United States: epidemiology and transmission to women. *Am J Public Health.* 1992;82:220–224.
2. Centers for Disease Control. U.S. AIDS cases reported through June, 1992. *HIV/AIDS Surveill.* July 1992:8(Table 3).
3. Doll L, Petersen L, White CR, Johnson ES, Ward JW, The HIV Blood Donor Study Group. Homosexually and non-homosexually identified men who have sex with men: a behavioral comparison. *Sex Res.* 1992;29:1–14.
4. Earl WL. Married men and same-sex activity: a field study on HIV risk among men who do not identify as gay or bisexual. *J Sex Marital Ther.* 1990;16:251–257.
5. Ames LJ, Beeker C, Rose DT. Continuing risk behavior among bisexually active men. Presented at the 118th Annual Meeting of the American Public Health Association; October 1990; New York, NY.
6. Beeker C, Rose DT, Ames LJ. Marginal men, mainstream risk. Poster presented at the Sixth International AIDS Conference; June 1990; San Francisco, Calif.
7. Handsfield HH, Wardell M, Goldstein L, Shriver K. Screening and diagnostic performance of enzyme immunoassay for antibody to lymphadenopathy-associated virus. *J Clin Microbiol.* 1987;25:879–884.
8. Taylor RN, Hearn TC, Schalla WO, Valdiserri RO. Indirect immunofluorescence test performance and questionnaire results from the Centers for Disease Control Model Performance Evaluation Program for human immunodeficiency virus, type 1 testing. *J Clin Microbiol.* 1990;28:1799–1807.
9. Centers for Disease Control. Interpretation and use of the Western blot assay for serodiagnosis of human immunodeficiency virus, type 1 infections. *MMWR.* 1989; 38(suppl 7):S-7.
10. Dean AG, Dean JA, Burton AH, Dicker RC. *Epi Info, Version 5: A Word processing, database, and statistics program for epidemiology on microcomputers.* Stone Mountain, Ga: USD Inc; 1990.
11. Lifson AR. Men who have sex with men: continued challenges for preventing HIV infection and AIDS. *Am J Public Health.* 1992;82:166–167.

Acculturation and Gender Differences in Sexual Attitudes and Behaviors: Hispanic vs Non-Hispanic White Unmarried Adults

Barbara VanOss Marín, PhD, Jeanne M. Tschann, PhD, Cynthia A. Gómez, PhD, and Susan M. Kegeles, PhD

ABSTRACT

Understanding how acculturation and gender affect Hispanics' sexual behavior is needed to prevent infection with the human immunodeficiency virus. We examined differences in and correlates of condom use among 398 Hispanics and 540 non-Hispanic Whites in San Francisco who were part of a random probability sample of unmarried adults. Hispanic women reported fewer sexual partners than all other groups. Condom use was low in all groups, but Spanish-speaking Hispanic women reported lower condom use than White women. Hispanics, generally, had poorer attitudes toward condoms and were less likely than non-Hispanic Whites to believe they could avoid acquired immunodeficiency syndrome. Hispanics need targeted prevention interventions. (*Am J Public Health.* 1993;83: 1759–1761)

Introduction

In the United States, acquired immunodeficiency syndrome (AIDS), the human immunodeficiency virus (HIV), and sexually transmitted diseases are disproportionately found in Hispanic populations.[1–4] Thus, models for understanding sexual risk behaviors among these populations are urgently needed.

Culturally appropriate programs must be based on a clear understanding of the antecedents of risk behaviors.[5,6] Among Hispanics, acculturation—a process of adaptation in which immigrants alter their attitudes and behaviors to more closely resemble those of the host society—is an important predictor of many health-related behaviors, including cigarette smoking, alcohol use, and early sexual initiation.[7–9] For certain behaviors, such as smoking and drinking, acculturation affects each gender differently, making women more likely to adopt risky behaviors but not men.[8,9] Given the traditional attitudes toward gender roles reported among less acculturated Hispanics,[10] gender and acculturation should be important factors in understanding sexual behavior among Hispanics.

The present study sought to identify ethnic, acculturative, and gender differences both in the number of sexual partners and in condom attitudes and behaviors among a random sample of young unmarried Hispanic and non-Hispanic White adults in San Francisco.

Methods

Respondents

In total, 1770 respondents were interviewed from 1988 to 1989 for a research project, the AIDS in Multiethnic Neigh-

Barbara VanOss Marín and Cynthia A. Gómez are with the Department of Epidemiology and Biostatistics and the Center for AIDS Prevention Studies, Institute of Health Policy Studies; Jeanne M. Tschann is with the Department of Psychiatry; and Susan M. Kegeles is with the Department of Medicine and the Center for AIDS Prevention Studies, Institute of Health Policy Studies, at the University of California in San Francisco.

Requests for reprints should be sent to Barbara VanOss Marín, PhD, Center for AIDS Prevention Studies, University of California, 74 New Montgomery St, Suite 600, San Francisco, CA 94105.

This paper was accepted April 9, 1993.

TABLE 1—Demographic Variables for Each Gender, Ethnicity, and Language Group

	Non-Hispanic Whites		English-Speaking Hispanics		Spanish-Speaking Hispanics	
	Men (n = 269)	Women (n = 271)	Men (n = 100)	Women (n = 103)	Men (n = 117)	Women (n = 78)
Mean age, y	29.6	30.1	27.9	28.2	27.1	31.6
Education, %						
Less than high school	6	5	16	21	36	53
High school graduate	15	10	26	25	39	22
Some college	37	31	49	36	22	22
College graduate	25	34	4	10	3	3
Postgraduate work	15	20	5	8	0	0
Mean income, in thousands	$16.8	$15.0	$13.2	$13.2	$ 9.6	$ 8.4

TABLE 2—Partner Type for Each Gender, Ethnicity, and Language Group

	Non-Hispanic Whites		English-Speaking Hispanics		Spanish-Speaking Hispanics	
	Men (n = 269)	Women (n = 271)	Men (n = 100)	Women (n = 103)	Men (n = 117)	Women (n = 78)
Partner type, %						
Not sexually active	11	11	21	34	19	39
Monogamous, not new	30	40	25	36	32	46
Monogamous, new	9	9	6	9	14	10
Multiple partners	50	40	48	21	35	5
Mean partner type score[a]	1.90	1.79	1.81	1.17[b]	1.66	0.82[b]

[a]Scores were derived from the following scale: not sexually active = 0; monogamous, not new = 1; monogamous, new = 2; and multiple partners = 3.
[b]Tukey post hoc comparisons show these groups to differ significantly from all other groups.

TABLE 3—Condom Use for Each Gender, Ethnicity, and Language Group among Sexually Active Respondents

	Non-Hispanic Whites		English-Speaking Hispanics		Spanish-Speaking Hispanics	
	Men (n = 241)	Women (n = 242)	Men (n = 79)	Women (n = 68)	Men (n = 95)	Women (n = 48)
Condom use, %						
Never	55	53	52	71	66	79
Sometimes	39	36	37	23	23	15
Always	5	11	11	6	11	6
Mean condom use score[a]	1.50	1.58	1.59	1.35	1.44	1.27[b]

[a]Scores were derived from the following scale: never = 1, sometimes = 2, and always = 3.
[b]Tukey post hoc comparisons show this group to differ significantly from all other groups.

borhoods Study. This paper reports on 938 respondents who reported no homosexual or bisexual behavior in the 12 months prior to the interview and who self-identified as either Hispanic (n = 398) or White (n = 540). Table 1 reports the demographics of this sample.

Procedure

A complete description of the rationale for target population selection, sampling methods, and participation rates for the research appears elsewhere.[11] Briefly, the target population consisted of currently unmarried men and women, aged 20 to 44, who lived in 16 census tracts of San Francisco. Each household had the same probability of being selected (1 in 6.2). All eligible respondents in the household were interviewed. The response rate was 64%.

The interview included a detailed assessment of sexual behavior in the previous 12 months, a variety of attitudinal measures, and demographic items. Acculturation was not measured directly, but the language spoken in the interview was used as a reasonable proxy.[12,13] The instrument was translated into Spanish using a back-translation procedure.[14]

Statistical Analysis

A series of 3×2 analyses of covariance were conducted, with ethnicity/language and gender as independent variables, and age, education, and income as covariates. The first analysis included all respondents and used partner type (not sexually active; monogamous, long term; monogamous, less than 1 year; and multiple partnered) as the dependent variable. Additional analyses of covariance were based on sexually active respondents only and used condom use (always, sometimes, or never with all partners), condom attitude, or beliefs about avoiding AIDS as the dependent variable.

Results

Ethnic/Language and Gender Differences in Sexual Behavior

Partner type. Overall, 18% (n = 165) of respondents were not sexually active in the past year, 35% (n = 324) were monogamous with no new partner, 10% (n = 90) were monogamous with a new partner, and 38% (n = 359) had multiple partners. Ethnicity/language (F [2, 915] = 16.37, $P < .001$), gender (F [1, 915] = 30.45, $P < .001$), and their interaction (F [2, 915] = 7.33, $P < .001$) are all significant predictors of partner type (see Table 2).

Condom use. Condom use was low: among the 773 sexually active respondents, 58% (n = 452) reported never using condoms, 33% (n = 255) reported sometimes using condoms, and 9% (n = 66) reported always using condoms. Only the interaction between ethnicity/language and gender is significant (F [2, 755] = 3.12, $P < .05$) (see Table 3).

Ethnic/Language and Gender Differences in Psychological Variables

Condom attitude. Women had more positive attitudes about condom use than men (F [1, 764] = 35.80, $P < .0001$), and non-Hispanic Whites had more positive attitudes than Spanish-speaking Hispanics (F [2, 764] = 4.43, $P < .02$). The interaction between ethnicity/language and gender is not significant (see Table 4).

Beliefs about avoiding AIDS. Spanish-speaking Hispanics had the lowest belief that they could do something to avoid AIDS (F [2, 764] = 68.54, $P < .0001$). Women had higher scores than men (F [1, 764] = 11.02, $P < .001$). The interaction between ethnicity/language and gender is not significant (see Table 4).

TABLE 4—Attitudinal Variables for Each Gender and Each Ethnicity and Language Group among Sexually Active Respondents

	Men (n = 415)	Women (n = 358)	Non-Hispanic Whites (n = 483)	English-Speaking Hispanics (n = 147)	Spanish-Speaking Hispanics (n = 143)
Condom attitude[a,b]	2.2	2.5	2.4	2.5	2.1
Beliefs about avoiding AIDS[a,c]	3.4	3.6	3.7	3.5	2.8

[a]Values correspond to mean scores.
[b]Condom attitude included the following: vaginal intercourse is enjoyable with condoms, condoms take the fun out of sex, sex doesn't feel as good with condoms, and condoms are physically uncomfortable to use. Range: 1 to 5, with higher scores indicating more positive attitudes.
[c]Beliefs about avoiding AIDS included the following: little can be done if one's partner does not want to use condoms, and little can be done to avoid AIDS. Range: 1 to 4, with higher scores representing stronger beliefs that one can avoid AIDS.

Discussion

Unmarried Hispanic women reported fewer partners than unmarried Hispanic men or unmarried non-Hispanic White men and women. On the other hand, more than one third of the Spanish-speaking Hispanic men and almost half of the English-speaking Hispanic men in this study reported having multiple partners. In traditional Hispanic culture, such "macho" behavior is expected of a man.[10,15,16] Accordingly, it is often erroneous for a Hispanic woman with only one partner to consider herself at low risk. Moreover, among sexually active respondents, Spanish-speaking Hispanic women had the lowest rates of condom use. Thus, Hispanic women can be seen as both more protected and more at risk; although they reported lower sexual activity, those who were sexually active had lower rates of condom use while their male counterparts reported high rates of multiple partners. These findings form part of a larger context in which Spanish-speaking women are less knowledgeable about HIV in general than non-Hispanic White women,[17] carry condoms less often,[18] and see condoms as unpleasureable.

In this study, men were more likely than women to view condoms as interfering with their sexual pleasure. Respondents were fairly negative about condoms, with men, and particularly Spanish-speaking men, being most negative. Both Spanish-speaking men and women believed there was little they could do to avoid AIDS. This suggests the need for special campaigns in Spanish to address these beliefs.

This random sample of unmarried adults drawn from 16 census tracts in San Francisco may be representative particularly of Hispanic and non-Hispanic Whites who are young, unmarried, and living in urban areas. However, generalization to Hispanic subgroups other than Mexicans and Central Americans should be done with caution.

AIDS-related attitudes and behaviors are strongly associated with gender, ethnic group, and level of acculturation. Traditional Hispanic culture may pose particular challenges to AIDS prevention educators, especially to those attempting to reach and influence less acculturated Hispanic women. □

Acknowledgments

This research was funded by center grant MH42459 from the National Institute of Mental Health and the National Institute on Drug Abuse, and grants MH46777 and MH46789 from the National Institute of Mental Health.

The authors wish to acknowledge Drs Mindy Thompson Fullilove, Stephen B. Hulley, and Joseph Catania, who initiated and directed the AIDS in Multiethnic Neighborhoods Study, as well as the staff at the Survey Research Center, who collected the interview data. Drs Olga Grinstead and Thomas Coates provided particularly helpful comments on an earlier draft of this paper.

A more detailed version of this paper is available from Barbara VanOss Marín.

References

1. Centers for Disease Control. HIV/AIDS surveillance. *HIV/AIDS Surveillance.* 1991;June:1–18.
2. Selik RM, Castro KG, Pappaioanou M. Centers for Disease Control Surveillance Summaries. *MMWR.* 1988;37:1–10.
3. Centers for Disease Control. Continuing increase in infectious syphilis—United States. *MMWR.* 1988;37:35–38.
4. Centers for Disease Control. Syphilis and congenital syphilis—United States, 1985–1988. *MMWR.* 1988;37:486–489.
5. Coates T, Greenblatt R. Behavioral change using community level interventions. In: Holmes K, ed. *Sexually Transmitted Diseases.* New York, NY: McGraw-Hill; 1989:1075–1080.
6. Coates T. Strategies for modifying sexual behavior for primary and secondary prevention of HIV disease. *J Consult Clin Psychol.* 1990;58:57–69.
7. Anashensel CS, Becerra RM, Fielder EP, Weiss NS. Onset of fertility-related events during adolescence: a prospective comparison of Mexican American and non-Hispanic White females. *Am J Public Health.* 1990;80:959–964.
8. Marín G, Pérez-Stable EJ, Marín BV. Cigarette smoking among San Francisco Hispanics: the role of acculturation and gender. *Am J Public Health.* 1989;79:196–199.
9. Caetano R. Patterns and problems of drinking among US Hispanics. In: *Report of the Secretary's Task Force on Black and Minority Health.* Washington, DC: US Dept of Health and Human Services; 1986;7:142–186.
10. Pavich EG. A Chicano perspective on Mexican culture and sexuality. *J Soc Work Hum Sexuality.* 1986;4:47–65.
11. Fullilove MT, Wiley J, Fullilove RE, et al. Risk for AIDS in multiethnic neighborhoods of San Francisco: the population-based AMEN study. *West J Med.* 1992;157:32–40.
12. Marín G, Sabogal F, Marín B, Otero-Sabogal R, Pérez-Stable EJ. Development of a short acculturation scale for Hispanics. *Hispanic J Behav Sci.* 1987;9:183–205.
13. Rogler LH, Cortes DE, Malgady RG. Acculturation and mental health status among Hispanics. *Am Psychol.* 1991;46:585–597.
14. Marín G, Marín B. *Research with Hispanic Populations.* Newbury Park, Calif: Sage Publications; 1991.
15. Burgos NM, Diaz Pérez YI. An exploration of human sexuality and the Puerto Rican culture. *J Soc Work Hum Sexuality.* 1986;4:135–150.
16. Comas-Diaz L. Mainland Puerto Rican women: a sociocultural approach. *J Community Psychol.* 1988;16:21–31.
17. Dawson D. *AIDS Knowledge and Attitudes of Hispanic Americans: Provisional Data from the National Health Interview Survey.* Washington, DC: National Center for Health Statistics; 1989.
18. Marín BV, Marín G. Predictors of condom accessibility among Hispanics in San Francisco. *Am J Public Health.* 1992;82:592–595.

Inappropriate Halsted Mastectomy and Patient Volume in Italian Hospitals

Roberto Grilli, MD, Franco Mainini, MD, Angelo Penna, MD, Guido Bertolini, MD, Nicola Scorpiglione, MD, Valter Torri, MD, and Alessandro Liberati, MD

ABSTRACT

To study whether Halsted mastectomy was used only when properly indicated, a prospective survey was undertaken on the process of care of 985 breast cancer patients seen consecutively at 62 general hospitals in Northern and Central Italy. Overall, 79% of Halsted mastectomies were performed inappropriately. The procedure was less likely to be performed on more educated patients and, other factors considered, on those seen at hospitals with larger volume. We conclude that the measurement of utilization of a surgical procedure for which only a few appropriate indications exist may help identify important relationships between hospital characteristics and quality of surgical care. (*Am J Public Health.* 1993;83:1762–1764)

Introduction

A few studies have indicated that care provided to cancer patients outside specialized institutions is often of lower quality, even in areas where the availability of a substantial body of scientific knowledge should have led to a general consensus about what constitutes appropriate care.[1–4] Caregivers' personal characteristics (age, education, training, experience, and skill), attitudes, and beliefs are increasingly seen as the preeminent cause when large variations in practice patterns exist.[5,6]

A mail survey carried out in 1987 to investigate physicians' opinions on breast cancer surgery in Italy showed, for example, that doctors working in hospitals with larger volumes and those favoring patient participation in treatment decisions were more likely to recommend less radical surgery.[6,7] Based on these findings, one would expect unnecessary radical surgical procedures to be more common in hospitals with low breast cancer patient volumes.

To investigate whether the relationship held true in actual practice, we analyzed data from a study on patterns of breast cancer care carried out in Italy in 1983 and 1984, using the performance of Halsted mastectomy as indicator of quality of care.[8]

Methods

Data Collection and Criteria for Assessment of Surgical Appropriateness

The methodology of data collection has been reported in detail elsewhere.[3] Briefly, information on 1262 breast cancer patients was prospectively collected in 62 general hospitals from 11 Italian regions over the period from March 1983 through April 1984. All patients consecutively admitted and diagnosed at participating centers were eligible for entry into the study. Halsted mastectomies were considered inappropriate when they were performed on patients without tumor fixation to the pectoral muscle and with nodal status classified as N_{0-1}.[9] We excluded from this analysis 277 patients for whom no specific recommendation was included in the guidelines because their tumors were adherent to the skin; they had clinical involvement of the supraclavicular axillary nodes; they had distant metastasis; they did not undergo surgery; or they lacked information on the clinical stage of disease.[9] Thus, the present report is based on data from 985 breast cancer patients.

Data Analysis

We first performed a series of univariate analyses to see whether the probability of being treated inappropriately with Halsted radical mastectomy was associated with patients' social and clinical characteristics, hospital size (number of beds), or volume of patients (based on the total number of patients enrolled in the study over the 1-year accrual period). The Mantel-Haenszel chi-square test and (when appropriate) the Mantel-Haenszel chi-square test for linear trend[10] were used to assess the statistical significance of the association. The strength of the associations was assessed by means of odds ratios (ORs) with 95% confidence intervals (CIs). Adjustment for potential confounding factors was done with the Mantel-Haenszel procedure.[10] Final analysis was performed through logistic regression[11]: the model included as independent variables all vari-

Roberto Grilli, Franco Mainini, Angelo Penna, Guido Bertolini, Valter Torri, and Alessandro Liberati are with the Laboratorio di Epidemiologia Clinica (Centro Catullo e Daniela Borgomainerio), Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy. Nicola Scorpiglione is with the Laboratorio di Epidemiologia e Farmacologia Clinica, Consorzio Mario Negri Sud, S. Maria Imbaro, Chieti, Italy.

Requests for reprints should be sent to Roberto Grilli, MD, Laboratorio di Epidemiologia Clinica, Istituto di Ricerche Farmacologiche Mario Negri, Via Eritrea 62, 20157 Milan, Italy.

This paper was accepted March 23, 1993.

TABLE 1—Social Characteristics of Breast Cancer Patients Enrolled in the Survey (n = 985) and Their Relationship to Inappropriate Use of the Halsted Mastectomy

	n	% Who Received Inappropriate Halsted Mastectomy	P
Age			
30–48 y	286	26	
49–58 y	250	30	
59–69 y	239	27	
>69 y	210	24	NS
Marital status			
Married	681	28	
Unmarried	302	25	NS
Occupation			
Unemployed	541	27	
Retired	106	26	
Blue- and white-collar	247	26	
Professional	91	29	NS
Education			
0–4 y	280	31	
5 y	440	26	
6–8 y	159	26	
>8 y	106	17	.016
OR[a] = 0.53; 95% CI = 0.31, 0.89			

Note. NS = not significant.
[a]The odds ratio (OR) compares patients with more than 8 y education vs those with ≤8 y education, with the latter as reference category, adjusting for age and stage of disease.

TABLE 2—Clinical Characteristics of Breast Cancer Patients Enrolled in the Survey (n = 985) and Their Relationship to Inappropriate Use of the Halsted Mastectomy

	n	% Who Received Inappropriate Halsted Mastectomy	P
Tumor size at clinical examination			
≤2 cm	292	23	
>2 cm[a]	551	36	<.001
OR = 0.54; 95% CI = 0.39, 0.75			
Nodal status at clinical examination			
$N_{0/1a}$	781	25	
N_{1b}[a]	204	33	.035
OR = 0.70; 95% CI = 0.50, 0.97			
Clinical stage of disease			
Stage III[a]	87	40	
Stage II	616	28	
Stage I	227	22	.001
OR[b] = 0.49; 95% CI = 0.31, 0.77			

[a]Reference category.
[b]The odds ratio (OR) compares stage I and II vs stage III patients, adjusting for age and years of education.

TABLE 3—Rates of Use of Inappropriate Halsted Radical Mastectomy in Patients Eligible for More Conservative Surgery, by Hospital Size and Patient Volume in Participating Centers

	1–20 Patients/y		>20 Patients/y	
	n	%	n	%
<500 beds				
Inappropriate Halsted	53	37	20	32
Other surgery	90	63	42	68
Total	143	14	62	6
OR = 1.24; 95% CI = 0.66, 2.32				
>500 beds				
Inappropriate Halsted	80	39	113	19
Other surgery	123	61	464	80
Total	203	21	577	59
OR = 2.67; 95% CI = 1.88, 3.78				

ables with an individually statistically significant ($P < .05$) association, identified with a stepwise procedure.

Results

The mean age of the patients whose cases were reviewed was 57 years (range 30–84). Overall, 366 patients (34%) were treated with Halsted radical mastectomy; 266 procedures (79%) were inappropriate. As shown in Table 1, more educated patients were 50% less likely to undergo Halsted mastectomy; the association held true even after adjusting for age and disease stage (OR = 0.53; 95% CI = 0.31, 0.89).

Halsted mastectomy was more frequent among patients with more advanced disease: 41% of clinical stage III patients, compared with 22% of stage I and 26% of stage II patients, had the procedure ($\chi^2 = 9.59$, $P = .001$) (Table 2). The same association held true both when pathologic stage was considered and when its two components (size of the primary tumor and nodal status) were considered separately (Table 2).

The use of inappropriate radical mastectomy was closely related to both volume of patients and hospital size. Centers with lower patient volume (1 through 10 and 11 through 20) had higher inappropriateness rates—43% and 36%, respectively, compared with 21% for hospitals seeing more than 20 patients per year (χ^2 for linear trend = 35.5, $P < .001$). A univariate analysis showed that patients seen at low-volume hospitals had double the probability of receiving a Halsted mastectomy (OR = 2.37; 95% CI = 1.78, 3.16). Hospital size also played a role: Halsted mastectomies were inappropriately performed on 36% of patients seen at smaller centers, compared with 24% cared for at larger hospitals (OR = 1.68; 95% CI = 1.21, 2.33).

Table 3 shows the proportion of patients treated with inappropriate Halsted mastectomy within the two categories of hospital size stratified by patient volume. Although the association between quality of surgery and patient volume had the same direction in the two strata, it was statistically significant only in the larger centers, where patients admitted to low-volume hospitals had double the probability of an unnecessary radical procedure (OR = 2.67; 95% CI = 1.88, 3.78).

A multivariate analysis showed that women seen at high-volume hospitals were less likely to be treated with an unnecessary radical mastectomy (OR = 0.36; 95% CI = 0.23, 0.55) (Table 4). Based on the logistic model including all the relevant covariates, probabilities of being treated with an inappropriate Halsted mastectomy were 48%, 43%, and 24% for hospitals recruiting 1 through 10, 11 through 20, and more than 20 patients per year, respectively.

Discussion

Approximately one third of the breast cancer patients in our study were treated with a surgical procedure that, by current standards, should be performed in only a minority of cases; it is thus not surprising that in almost 80% of the cases the procedure was judged to be inappropriate. It is hard to believe that hospital organization and availability of specific facilities account for these findings. If these factors play any role it should be only in the diffusion of conservative procedures, for

TABLE 4—Logistic Regression Model Assessing the Association between Hospitals' Annual Patient Volume and Use of Inappropriate Halsted Mastectomy, with All Potential Confounding Factors Simultaneously Controlled

Covariates	Odds Ratio	95% CI	P
Tumor size[a]			
≤2 cm	0.60	0.43, 0.84	.004
Nodal status[b]			
$N_{0/1/a}$	0.67	0.46, 0.98	.04
Education[c]			
≥8 y	0.54	0.31, 0.94	.03
Hospital's annual patient volume[d]			
11–20	0.82	0.51, 1.32	.42
>20	0.36	0.23, 0.55	.001

[a]Reference category, >2 cm.
[b]Reference category, N_{2b}.
[c]Reference category, <8 y.
[d]Reference category, 1–10.

TABLE 5—Probability of Being Treated with an Inappropriate Halsted Mastectomy, by Center's Annual Patient Volume, in 1983/84 and 1986/87

Annual Patient Volume	1983/84		1986/87	
	OR	95% CI	OR	95% CI
11–20	0.78	0.31, 1.83	0.33	0.14, 0.79
>20	0.35	0.16, 0.77	0.19	0.10, 0.48

Note. The reference category is 1–10 patients per year.

which postsurgical radiotherapy is mandatory.

The key finding of this study—the relationship between patient volume and quality of care—is not new and is consistent with the findings of several other studies. Moreover, it held true for other aspects of breast cancer care, such as timeliness of diagnosis and performance of radiotherapy after conservative surgery.[12–14] The latter fact is of specific importance because it offers an opportunity to identify medical communities at risk of providing poor quality care and to identify sites more in need of educational interventions.

Although one might argue that our findings are based on outdated practice patterns, an analysis based on information gathered in a more recent study, at the end of the 1980s, showed that the use of Halsted mastectomy was still high.[15] Moreover, even in the more recent study we found the same association between patient volume and use of unnecessary radical procedures. As Table 5 shows, data from both surveys are consistent with a 70% reduction in the risk of being treated with an inappropriate Halsted mastectomy in high-volume centers.

A few potential limitations of our study should be discussed. Our measure of volume was only a proxy for the actual number of patients seen in 1 year (see Methods). However, as the study was descriptive, it is hard to see why patients should have been selectively excluded depending upon the surgery they had had. Moreover, it is likely that the association we found would have been stronger had we used a more precise measure of the actual number of patients treated.

Some hospital self-selection may have occurred, given that participation was voluntary. A closer comparison of our hospital sample with the universe of Italian hospitals[16] indicated, however, that proportionally larger and better technically equipped hospitals entered this study, making our results even more worrying. Hospitals seeing more than 20 breast cancer patients per year in Italy account for approximately one sixth of the estimated 30 000 new cases of the disease diagnosed every year in the country.[16] Therefore, we estimate that about 2500 avoidable (inappropriate) Halsted procedures were carried out during the study period. □

Acknowledgments

This study was supported by Progetto Finalizzato "Oncologia," Consiglio Nazionale delle Ricerche, grant 88.0095.44. Dr Bertolini was the recipient of a fellowship from the Banca Popolare di Milano; Dr Scorpiglione was the recipient of a fellowship from the Centro di Formazione e Studi per il Mezzogiorno—FORMEZ (Progetto Speciale "Ricerca Scientifica e Applicata nel Mezzogiorno").

We thank the anonymous referees for their useful suggestions in the revision of this manuscript.

References

1. Liberati A, Andreani A, Colombo F, Confalonieri C, Tognoni G. Care of cancer patients in thirty-one Italian general hospitals: methodological aspects and general findings. *Eur J Cancer Clin Oncol.* 1983;19:179–185.
2. Hanks GE, Kramer S, Diamond JJ, Herring DF. Patterns of care outcome survey: national outcome for six disease sites. *Am J Clin Oncol.* 1982;5:349–353.
3. GIVIO. Diagnosis and first-line treatment of breast cancer in Italian general hospitals. *Tumori.* 1986;72:273–283.
4. Interdisciplinary Group for Cancer Care Evaluation (GIVIO). Diagnosis and first-line treatment of patients with lung cancer in Italian general hospitals. *Tumori.* 1989;75:163–167.
5. Liberati A, Patterns WB, Biener L, McNeil BJ. Determinants of physicians' preferences for alternative treatments in women with early breast cancer. *Tumori.* 1987;73:601–609.
6. GIVIO. Survey of treatment of primary breast cancer in Italy. *Br J Cancer.* 1988;57:630–634.
7. Liberati A, Apolone G, Nicolucci A, et al. The role of attitudes, beliefs, and personal characteristics of Italian physicians in the surgical treatment of early breast cancer. *Am J Public Health.* 1991;81:38–42.
8. Wickerham DL, Fisher B. Surgical treatment of primary breast cancer. *Semin Surg Oncol.* 1988;4:226–233.
9. Forza Operativa Nazionale sul Carcinoma Mammario. I tumori della mammella. Protocollo di trattamento. CNR Progetto Finalizzato. Controllo della Crescita Neoplastica. Milano, Italia, Aprile 1982.
10. Schlesselman JJ. *Case-Control Studies: Design, Conduct, Analysis.* New York, NY: Oxford University Press; 1982.
11. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods.* New York, NY: Van Nostrand Reinhold; 1982.
12. Flood AB, Scott RS, Ewy W. Does practice make perfect? part I: the relationship between hospital volume and outcomes for selected diagnostic categories. *Med Care.* 1984;22:98–114.
13. Flood AB, Scott RS, Ewy W. Does practice make perfect? part II: the relationship between volume and outcomes and other hospital characteristics. *Med Care.* 1984;22:115–125.
14. Hughes RG, Hunt SS, Luft HS. Effects of surgeon volume and hospital volume on quality of care in hospitals. *Med Care.* 1987;25:489–503.
15. Grilli R, Apolone G, Marsoni S, et al. The impact of patient management guidelines on the care of breast, colorectal and ovarian cancer patients in Italy. *Med Care.* 1991;29:50–63.
16. Progetto Oncologia Femminile. L'organizzazione dell' assistenza in oncologia ginecologia. *Nuovi Argomenti in Medicina.* 1991;9:400–407.

Predictors of Nursing Home Admission in a Biracial Population

Marcel E. Salive, MD, MPH, Karen S. Collins, MD, MPH, Daniel J. Foley, MS, and Linda K. George, PhD

ABSTRACT

Racial differences in predictors of institutionalization were studied in a biracial North Carolina cohort (n = 4074). During 3 years of follow-up, 8.5% of Whites and 6.4% of African Americans were admitted to nursing homes. African Americans were one half as likely as Whites to be institutionalized after adjustment for other risk factors. Among Whites, impaired activities of daily living and cognition were the strongest predictors; among African Americans, impaired instrumental activities of daily living and prior history of nursing home use were strongest. Racial differences in nursing home use were not explained by financial and social support or physical and cognitive impairment. (*Am J Public Health.* 1993;83:1765–1767)

Introduction

Older adults seek nursing home care primarily because of physical or mental impairment secondary to illness in advanced age and the lack of informal care.[1–3] Racial disparities have been reported such that 27% of African Americans can expect to use nursing homes in their lifetime, as compared with 38% of Whites.[4]

This analysis was designed to (1) compare nursing home use by African Americans and Whites while controlling for social support, economic status, and physical and cognitive function and (2) assess the relative importance of selected predictors of institutionalization within each racial group.

Methods

Study Population

The data were derived from the five-county North Carolina community of the Established Populations for Epidemiologic Studies of the Elderly, a collaborative longitudinal study of men and women 65 years of age or older. Details of the methods have been published.[5,6] During 1986 and 1987, trained interviewers conducted household surveys to collect baseline information from 4163 people (80% of eligible respondents). This study includes 4074 community-dwelling persons, classified by the interviewer as Black (55%) or White, who were followed for up to 3 years. Other races and persons without adequate follow-up to assess subsequent institutionalization were excluded.

Annual interviews ascertained interval admission to intermediate or skilled nursing facilities and personal care homes. Follow-up rates were greater than 95% each year among those initially interviewed. If a participant died, institutionalization since the last interview was ascertained from a proxy.

During the initial interview, information on age, gender, education, household income, marital status, household composition, types of medical insurance, urban or rural residence, and prior use of health services (home health aide, visiting nurse, or nursing home) was obtained. Social support was measured as perceived availability of support and instrumental support.[7] Physical function was measured as limitations in activities of daily living[8,9] and in instrumental activities of daily living.[10] Cognitive impairment was defined as four or more errors on the modified Short Portable Mental Status Questionnaire.[11,12]

Statistical Analysis

The baseline population was used to directly adjust the incidence of nursing home admission by age.[13] Logistic regression models used nursing home admission as the dependent variable in multivariate analyses examining the hypothesis that racial differences could be explained by demographic characteristics and social and functional status.[14,15] Logistic regression models stratified by race were examined to evaluate whether the factors associated with institutionalization differed by race. Sampling weights were used in all analyses to represent population estimates for the five counties.

Results

During the follow-up, 311 (7.6%) persons reported at least one nursing home admission. Blacks were institutionalized 24% less than Whites (3-year age-adjusted incidence rates of 6.4% and 8.5%, respectively; Table 1). Among those 85 years of age or older, Black women had the lowest rates.

Marcel E. Salive, Karen S. Collins, and Daniel J. Foley are with the Epidemiology, Demography and Biometry Program, National Institute on Aging, Bethesda, Md. Karen S. Collins is also with the General Preventive Medicine Residency Program, Johns Hopkins University School of Hygiene and Public Health, Baltimore, Md. Linda K. George is with the Center for the Study of Aging and Human Development, Duke University, Durham, NC.

Requests for reprints should be sent to Marcel E. Salive, MD, MPH, Epidemiology, Demography and Biometry Program, National Institute on Aging, National Institutes of Health, 7201 Wisconsin Ave, Gateway Bldg, Suite 3C309, Bethesda, MD 20892.

This paper was accepted April 26, 1993.

TABLE 1—Nursing Home Use According to Age, Race, and Gender: North Carolina Established Populations for Epidemiologic Studies of the Elderly, 1986 through 1990

Race/Age Group, y	No.	Use Since Baseline,[a] %	Lifetime History of Use,[a] %		
			Survivors	Deceased	Total
White men					
65–74	458	1.7	2.8	5.3	3.2
75–84	152	12.3	7.9	26.9	13.5
85+	32	31.3	30.0	45.5	36.6
White women					
65–74	670	4.1	3.8	13.0	4.4
75–84	426	15.4	16.4	33.5	18.6
85+	100	30.0	35.0	40.1	37.0
Total[b]	1838	8.5	8.5	17.1	10.2
Black men					
65–74	531	3.8	3.9	11.1	5.1
75–84	201	8.2	6.5	16.1	8.9
85+	46	27.8	18.2	41.9	27.8
Black women					
65–74	831	3.6	3.0	15.2	4.0
75–84	488	9.5	7.5	26.1	10.3
85+	139	16.0	11.6	29.3	17.0
Total[b]	2236	6.4	4.9	17.1	7.2

[a]Weighted to account for sampling.
[b]Age adjusted to the population of the community at baseline.

TABLE 2—Predictors of Nursing Home Admission According to Race: North Carolina Established Populations for Epidemiologic Studies of the Elderly, 1986 through 1990

Predictor	Whites (n = 1838)			Blacks (n = 2236)		
	Participants with Characteristic, %	Odds Ratio	95% CI	Participants with Characteristic, %	Odds Ratio	95% CI
Married	54.8	0.75	0.3, 1.7	44.1	0.56	0.3, 1.1
Lives alone	30.4	2.0	1.2, 3.6	26.8	1.2	0.7, 2.1
Children (per child)	85.5	0.84	0.75, 0.95	84.0	0.94	0.88, 1.01
High perceived availability of social support	85.0	0.80	0.66, 0.95	75.0	0.82	0.72, 0.93
Income <$5000/y	16.2	1.6	0.99, 2.50	42.1	1.3	0.8, 2.1
Medicaid enrollment	3.1	2.0	0.6, 5.5	10.0	1.8	0.96, 3.40
Urban residence	57.0	1.8	1.1, 2.8	56.0	1.3	0.8, 2.0
Physical function						
1–2 ADL impairments	7.2	1.7	0.9, 3.3	7.7	1.2	0.7, 2.2
≥3 ADL impairments	3.1	5.1	2.0, 13.0	4.4	1.7	0.8, 3.4
Instrumental ADL (per impairment)	26.7	1.0	0.9, 1.2	36.6	1.3	1.2, 1.5
Cognitive impairment	6.3	4.7	2.2, 10.0	19.1	1.2	0.7, 2.0
Prior nursing home use	2.7	1.8	0.7, 4.6	1.4	5.2	1.7, 16.0

Note. ADL = activities of daily living; CI = confidence interval. Relative odds were computed from race-specific multivariate logistic regression models based on nursing home admissions of 171 White participants ($r = .48$, predictive accuracy coefficient = .85) and 140 Black participants ($r = .42$, predictive accuracy coefficient = .79). Models were adjusted for age, sex, and all main effects shown.

Based on a model for the entire cohort, the following characteristics were significantly ($P < .05$) associated with nursing home use: age, White race, impairment in activities of daily living, cognitive impairment, Medicaid eligibility, urban residence, prior nursing home use, living alone, unmarried, low perceived availability of support, and having fewer children. Blacks were only half as likely as Whites to be institutionalized (relative odds = 0.50, 95% CI = 0.35, 0.72) after these variables were controlled.

Among elderly Whites, urban residence, living alone, low perceived social support, impairment in activities of daily living, cognitive impairment, and having fewer children were significantly ($P < .05$) associated with nursing home use (Table 2). Among Whites, limitations in three or more activities of daily living and cognitive impairment increased the risk of institutionalization approximately fivefold. Among African Americans, nursing home admission was significantly associated with fewer predictors, specifically prior nursing home use, impairment in instrumental activities of daily living, and perceived social support.

Discussion

This longitudinal analysis of nursing home use in a biracial North Carolina community serves to clarify and challenge some long-standing beliefs about explanations for racial differences in institutionalization. In 3 years of follow-up, African Americans were admitted to nursing homes about half as often as Whites after adjustment for several confounding factors. Although this population is geographically based, these differences have been reported nationally,[1] and the explanation may, in part, cut across regional lines.[16]

Our results highlight specific differences in predictors of institutionalization between these two racial groups. Among Whites, urban residents were more likely to be institutionalized than were their rural neighbors. Furthermore, physical and cognitive impairment were leading risk factors for admission of Whites but not African Americans. This may reflect differing cultural attitudes toward disability, a need to have the person at home regardless of impairment,[17,18] or limited knowledge of the long-term care system. Moreover, the Short Portable Mental Status Questionnaire may have divergent ability to accurately screen for impairment across racial groups. Although research in this area is limited, lifetime experiences such as segregated schools and job discrimination may contribute to poor performance on mental status questionnaires, leading to unreliability of such screening tests across cultures or ethnic groups.[19] Evaluation of more sensitive and specific screening instruments in the appropriate population is needed to address this issue. In African Americans, instrumental activities of daily living and prior nursing home

use may also reflect underlying physical and cognitive impairment.

We found significant unexplained differences in institutionalization rates between African Americans and Whites in this population. Therefore, either we were not able to capture one or more additional factors that influence this difference at the individual or community level or the measures that we used as confounders were incomplete. The baseline data collection was extensive but did not include personal or caregiver attitudes toward institutionalization. Also, we did not measure precisely the use of home care as an alternative for institutionalization.

These results suggest that, in addition to physical and cognitive changes that occur with aging, underlying social and historical differences may, at some level, influence current patterns of nursing home use. This analysis cannot address whether African Americans at home are better or worse off than their White counterparts. Our results do question, however, the way in which racial differences have been explained in the past without complete information. Future work should examine the current and future needs of African Americans and other minorities for access to long-term care services. □

Acknowledgments

This study was supported in part by contract N01-AG-1-2102 from the National Institute on Aging.

References

1. National Center for Health Statistics, Hing E, Sekscenski E, Strahan G. The National Nursing Home Survey; 1985 summary for the United States. *Vital Health Stat [13].* 1989;97. DHHS publication PHS 89-1758.
2. Foley D. Epidemiologic research on nursing home populations. In: Wallace RB, Woolson RF, eds. *The Epidemiologic Study of the Elderly.* New York, NY: Oxford University Press; 1992:234–241.
3. George LK, Maddox GL. Social and behavioral aspects of institutional care. In: Ory MG, Bond R, eds. *Aging and Health Care: Social Science and Policy Perspectives.* New York, NY: Routledge; 1989: 116–141.
4. Kemper P, Murtaugh CM. Lifetime use of nursing home care. *N Engl J Med.* 1991; 324:595–629.
5. Cornoni-Huntley J, Blazer DG, Lafferty ME, Everett DF, Brock DB, Farmer ME, eds. *Established Populations for Epidemiologic Studies of the Elderly: Volume II: Resource Data Book.* Bethesda, Md: National Institutes of Health, National Institute on Aging; 1990. NIH publication 90-495.
6. Foley DJ, Ostfeld AM, Branch LG, Wallace RB, McGloin J, Cornoni-Huntley JC. The risk of nursing home admission in three communities. *J Aging Health.* 1992;4:155–173.
7. Landerman R, George LK, Campbell RT, Blazer DG. Alternative models of the stress buffering hypothesis. *Am J Community Psychol.* 1989;17:625–642.
8. Katz S, Ford AB, Moskovitz AW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychological function. *JAMA.* 1963;185:914–919.
9. Branch LG, Katz S, Kniepman K, Papsidero JA. A prospective study of functional status among community elders. *Am J Public Health.* 1984;74:266–268.
10. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist.* 1969;9:179–186.
11. Kahn RL, Goldfarb AI, Pollack M, Peck A. Brief objective measures for the determination of mental status in the aged. *Am J Psychiatry.* 1960;117:326–328.
12. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc.* 1975;23:433–441.
13. Lilienfeld AM, Lilienfeld DE. *Foundations of Epidemiology.* 2nd ed. New York, NY: Oxford University Press; 1980.
14. Hastings RP, ed. *SUGI Supplemental Library User's Guide, Version 5 Edition.* Cary, NC: SAS Institute Inc; 1986.
15. Shah BV, Folsom RE, Harrell FE, Dillard CN. *RTILOGIT: Procedure for Logistic Regression on Survey Data.* Research Triangle Park, NC: Research Triangle Institute; 1987.
16. Weissert WG, Cready CM. Toward a model for improved targeting of aged at risk of institutionalization. *Health Services Res.* 1989;24:485–510.
17. Belgrave LL, Wykle M, Choi JM. Theoretical approaches to the use of institutionalization by African Americans. Presented at the annual meeting of the American Sociological Association, August 1991, Cincinnati, Ohio.
18. Mutran E. Intergenerational family support among Blacks and Whites: response to culture or to socioeconomic differences. *J Gerontol.* 1985;40:382–389.
19. Baker FM. Dementing illness and Black Americans. In: Jackson JS, ed. *The Black American Elderly: Research on Physical and Psychosocial Health.* New York, NY: Springer; 1988.

The Behavioral Risk Factor Surveillance System Questionnaire: Its Reliability in a Statewide Sample

Aryeh D. Stein, PhD, Ruth I. Lederman, MPH, and Steven Shea, MD

ABSTRACT

The reliability of the Behavioral Risk Factor Surveillance System questionnaire was assessed in a random sample of adults (n = 122) and a separate sample of Black and Hispanic adults (n = 200) in Massachusetts. The questionnaire was administered twice, 21 to 44 days apart, by telephone (210 completed reinterviews, 65% response rate for second administration). There were no statistically significant differences in the distribution of demographic or risk factor variables across administrations. Individual-level reliability (kappa for categorical variables, correlation for continuous variables) for demographic characteristics was more than 0.80 for White respondents and more than 0.60 for Black and Hispanic respondents. Employment and income were reported less consistently than other variables. Reliability coefficients for behavioral risk factors were generally above 0.70. Exceptions were variables with extreme distributions. These data support the use of the Behavioral Risk Factor Surveillance System questionnaire for surveillance and research. (*Am J Public Health.* 1993;83: 1768–1772)

Introduction

The Behavioral Risk Factor Surveillance System (BRFSS), an ongoing surveillance program developed and funded by the Centers for Disease Control and Prevention (CDC), is designed to estimate the prevalence of risk factors for the major causes of death in the United States.[1] Data from the BRFSS are a central component of federal and state activities designed to monitor progress toward achieving the health objectives for the year 2000.[2] We assessed the test–retest reliability of the BRFSS questionnaire in a representative sample of Massachusetts residents.

Methods

The BRFSS questionnaire has been described in detail elsewhere.[1] Demographic and risk factor variables are shown in Tables 1 and 2.

Two separate populations were recruited. The first consisted of respondents to the regular March 1992 sample of the Massachusetts BRFSS. The sampling methods were consistent with published CDC procedures. All interviews were conducted in English. Inability of the selected individual to participate in a telephone interview or speak English was considered a cause of ineligibility. With known eligible households as the denominator, the response rate for this survey was 65%.

The second population consisted of a separately recruited sample of Black and Hispanic residents of Massachusetts. The sampling frame for this study consisted of telephone exchanges in towns identified in the 1990 census as containing high concentrations of Blacks and Hispanics. Households were eligible if at least one household member was identified by the initial respondent as being Black or Hispanic. A random member of the household, irrespective of ethnicity, was interviewed. If necessary, a translation prepared by the CDC was used to conduct the interview in Spanish. With known eligible households as the denominator, the response rate for this survey was 39%.

A computer-assisted telephone interview system was used to collect data. In the first interview, respondents were not told that they might be contacted again. Callbacks for the second interview began 21 days after the first interview. Efforts were made to reinterview all of the original respondents, and these efforts were of the same intensity as those used to obtain the first interview. Eighty-six of 122 respondents to the regular survey (70%) and 124 of 200 respondents to the minority survey (62%) were successfully reinterviewed. Ninety-five percent of the 210 successful second interviews were completed within 44 days, and all were completed within 64 days. Univariate distributions of demographic variables, as recorded during the first interview, did not differ significantly ($P > .05$ for each demographic variable) between the 210 respondents and the 112 nonrespondents to the repeat interview.

Subjects were asked to identify race and whether they were of Hispanic origin. Subjects were classified as White non-Hispanic, Black non-Hispanic, or Hispanic according to the categories used by the CDC in reporting BRFSS data. Of 86 participants in the first sample, 11 were classified as non-White based on their responses to the first interview. Of 124 participants in the second sample, 15 were classified at the first interview as neither Black non-Hispanic nor Hispanic. Data from these 26 individuals were included in pooled analyses but were excluded from race/ethnicity-specific analyses.

At the time of the study, Aryeh D. Stein and Ruth I. Lederman were with the Bureau of Health Statistics, Research and Evaluation, Massachusetts Department of Public Health, Boston. Aryeh D. Stein was also with the Division of Epidemiology, Columbia University School of Public Health, New York, NY. Steven Shea was with the Division of Epidemiology, Columbia University School of Public Health, and the Division of Medicine, College of Physicians and Surgeons, Columbia University.

Requests for reprints should be sent to Aryeh D. Stein, PhD, Program in Epidemiology, A214 E Fee Hall, Michigan State University, East Lansing, MI 48824-1316.

This paper was accepted May 11, 1993.

Algorithms developed at the CDC (Dr Emma Frazier, written communication, July 1992) were used in computations of behavioral risk status from questionnaire items. Individuals with unknown values for demographic characteristics on either administration of the questionnaire were excluded from determination of group distributions for that administration. An exception was income; more than 5% of respondents refused to state or reported not knowing their annual household income and were considered a stratum for analysis. Individuals of unknown risk factor status were considered not to be at risk. This approach consistently underestimates prevalence because unknown responses are retained in the denominator.

The *z* test (for dichotomous variables and proportions, such as gender and risk status), the chi-square test (for polychotomous variables, such as educational attainment), and the *t* test (for means of continuous variables, such as cholesterol level) were used in assessing the consistency of distributions. The kappa statistic[3] (for dichotomous variables), unweighted kappa (for polychotomous variables), and Pearson's correlation coefficient (for continuous variables) were used in examining the consistency of individual responses. Individuals with unknown values on ei-

TABLE 1—Demographic Characteristics Available for Analysis in the 1992 Behavioral Risk Factor Surveillance System Questionnaire

Demographic Characteristic	Categories Used in Analysis
Age	Recorded in exact y
Sex	Male Female
Education	No high school/some high school High school graduate/some technical school/technical school graduate Some college but did not graduate College graduate/postgraduate or professional degree
Marital status	Married/living as a couple Divorced/separated Widowed Never married
Employment	Employed Self-employed Unemployed less than 1 y/unemployed more than 1 y Homemaker Retired Student
Income	Under $10 000/$10 000–$15 000 $15 000–$20 000/$20 000–$25 000/$25 000–$35 000 $35 000–$50 000/over $50 000 Unknown/refused
Race/ethnicity[a]	White but not of Hispanic origin Black but not of Hispanic origin Hispanic of any race Asian, Pacific Islander, or Native American/other/unknown/refused

Note. Categories separated by a slash were differentiated during the interview but combined for this report.
[a]Categories were defined on the basis of responses to the questions "What is your race?" and "Are you of Hispanic origin?"

TABLE 2—Risk Behaviors and Characteristics Available for Analysis in the Core Modules of the 1992 Behavioral Risk Factor Surveillance System Questionnaire

Module	Risk Behavior/Characteristic	Definition
Safety belts	Safety belt nonuse	Does not always use safety belt when driving or riding in a car
Hypertension	Recent blood pressure check	Has had blood pressure measured in past 2 y
	Diagnosed hypertensive	Has been told on more than one occasion that blood pressure was high or is taking blood pressure medication
Cholesterol	Recent cholesterol check	Has had cholesterol checked in past 5 y
	Cholesterol awareness	Can state own cholesterol level as a number
	Serum cholesterol	Cholesterol level stated as a number
	Hypercholesterolemia	Cholesterol level 200 mg/dL or higher
Exercise	Sedentary life-style	Did not exercise in past mo
	Regular aerobic exercise	Performed an aerobic activity at least three times per wk for at least 20 min per occasion in past mo
Smoking	Current cigarette use	Smoked 100 or more cigarettes over lifetime and currently smokes
	Cigarette consumption	Number of cigarettes smoked daily
	Former cigarette use	Smoked 100 or more cigarettes over lifetime but does not currently smoke
Alcohol	Any consumption	Any alcoholic drinks consumed in past mo
	Total consumption	Number of alcoholic drinks consumed in past mo
	Chronic consumption	Drank 60 or more alcoholic drinks in past mo
	Acute consumption	Drank five or more alcoholic drinks on an occasion at least once in past mo
	Drives after alcohol consumption	Drove a car after consuming too much alcohol at least once in past mo
Obesity	Body mass index	Weight in kilograms divided by square of height in meters
	Obese	Body mass index exceeding 27.8 kg/m^2 (men) or 27.3 kg/m^2 (women)
	Weight control	Currently trying to lose weight
Diabetes	Diagnosed diabetes	Doctor-diagnosed diabetes

TABLE 3—Consistency of Distribution of Demographic Characteristics across Two Administrations of the Behavioral Risk Factor Surveillance System Questionnaire among 210 Respondents: Massachusetts, 1992

Demographic Characteristic	All Respondents (n = 210)		White Non-Hispanic (n = 76)		Black Non-Hispanic (n = 64)		Hispanic (n = 45)	
	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
Mean age, y (SD)	40.7 (16.7)	40.4 (16.7)	46.7 (17.5)	46.4 (17.4)	40.7 (17.4)	40.7 (17.4)	34.0 (11.8)	33.4 (12.2)
Male sex, %	41.4	41.4	56.0	56.0	35.9	35.9	31.1	31.1
Education, %								
Less than high school	23.9	20.9	14.7	14.7	20.6	15.9	44.4	39.5
High school graduate	32.1	33.5	30.7	28.0	33.3	42.9	33.3	34.9
Some college	17.2	19.9	14.7	16.0	27.0	27.0	6.7	11.6
College graduate/professional	26.8	25.7	40.0	41.3	19.0	14.3	15.6	14.0
Marital status, %								
Married/live as couple	45.2	45.2	54.7	56.0	32.8	32.8	53.3	48.9
Divorced/separated	15.7	15.2	10.7	12.0	17.2	18.7	17.8	15.6
Widowed	5.2	5.2	8.0	6.7	6.3	6.3	2.2	4.4
Never married	33.8	34.3	28.7	25.3	43.7	42.2	28.7	31.1
Employment, %								
Employed	56.7	56.7	60.0	58.7	51.6	54.7	60.0	55.6
Self-employed	5.2	5.2	6.7	6.7	3.1	1.6	0.0	2.2
Unemployed	14.8	15.7	8.7	9.3	23.4	18.7	17.8	20.0
Homemaker	7.6	7.1	4.0	5.3	7.8	7.8	15.6	13.3
Retired	11.4	11.4	20.0	17.3	12.5	15.6	0.0	0.0
Student	4.3	3.8	2.7	2.7	1.6	1.6	6.7	6.9
Income, %								
<$15 000	30.9	32.9	20.0	20.0	26.6	37.5	48.9	40.0
$15 000–$35 000	32.9	30.9	30.7	29.3	39.1	31.3	26.7	31.1
>$35 000	29.5	30.5	45.3	48.0	23.4	25.0	17.8	17.8
Unknown/refused	6.7	5.7	4.0	2.7	10.9	6.3	6.7	11.1
Race/ethnicity, %								
White	39.1	37.6	...	...	...	...	...	...
Black	31.4	32.4	...	...	...	...	...	...
Hispanic	24.3	24.8	...	...	...	...	...	...
Other	5.2	5.2	...	...	...	...	...	...

Note. The total for all respondents is greater than the sum of White non-Hispanic, Black non-Hispanic, and Hispanic respondents as a result of exclusion of 28 subjects from ethnicity-specific analyses (see text for details). None of the differences between time 1 and time 2 values were significant ($P > .05$) within the whole sample or each ethnicity separately. The consistency of distribution of race/ethnicity was not compared for ethnicity-specific strata; it should be noted that the distribution of race/ethnicity in this sample does not reflect the ethnic composition of Massachusetts because of oversampling of Black and Hispanic residents.

ther administration of the interview for demographic characteristics or continuous variables and those for whom the continuous variable was not asked (e.g., number of cigarettes smoked among nonsmokers) were excluded from tests of individual consistency.

Results

Differences across administrations for demographic characteristics and risk factors were generally small (Tables 3 and 4). At the individual level, employment and income were reported less consistently than the other demographic variables (Table 5). Black and Hispanic respondents tended to report educational attainment, employment status, and income less consistently than did White respondents. Reliability was generally high for behavioral risk factors (Table 6), with most coefficients above .70 and many above .80. Consistency tended to be lower among Hispanic respondents than among White or Black respondents, although the differences were not marked.

Discussion

This paper suggests good test–retest reliability of the BRFSS questionnaire. Our data are consistent with the results of an earlier, more restricted study,[4] and extend these results to an analysis of the complete core BRFSS questionnaire in a statewide sample. Two studies have shown fair to good consistency of group mean prevalence of selected risk factors,[5,6] but no published studies have addressed the validity of individual responses. Our study should be relevant to all agencies participating in the BRFSS, because we adhered to BRFSS subject recruitment and data collection protocols. Our results concerning individual-level reliability are of potential interest to epidemiologic studies involving telephone-based surveys.[7]

True change between administrations of the questionnaire leads to lower estimates of reliability. Conversely, if the time period between administrations is too short, the questionnaires cease to be independent. In the present study, the time interval between the two administrations was 21 to 44 days. For most behaviors and characteristics, this interval was probably adequate. Questions concerning alcohol consumption and exercise, which addressed behaviors occurring in the previous month, showed the lowest consistency. For these variables, under the conditions of our study, it is not possible to separate questionnaire effects from true change. The sample size was too small to

TABLE 4—Consistency of Distribution of Risk Behaviors and Characteristics across Two Administrations of the Behavioral Risk Factor Surveillance System Questionnaire among 210 Respondents: Massachusetts, 1992

Module	Risk Behavior/Characteristic	All Respondents (n = 210)		White Non-Hispanic (n = 75)		Black Non-Hispanic (n = 64)		Hispanic (n = 45)	
		Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
Safety belts	Safety belt nonuse, %	60.5	61.4	52.0	58.7	62.5	57.8	66.7	68.9
Hypertension	Recent blood pressure check, %	96.7	95.7	96.0	94.7	96.9	96.9	95.6	93.3
	Diagnosed hypertension, %	16.7	13.8	16.0	16.0	26.6	21.9	13.3	6.7
Cholesterol	Recent cholesterol check, %	69.1	68.6	73.3	74.7	60.9	65.6	68.9	57.8
	Cholesterol awareness, no. (%)	50 (23.8)	49 (23.3)	26 (34.7)	30 (40.0)	5 (7.8)	4 (6.3)	9 (20.0)	6 (13.3)
	Cholesterol level (among those aware of level), mean (SD)	192.5 (41.5)	192.0 (59.3)	194.5 (42.6)	206.0 (65.0)	194.2 (60.7)	143.5 (58.3)	198.9 (41.9)	170.2 (41.3)
	Hypercholesterolemia, %	15.7	16.7	21.3	22.7	12.5	12.5	8.9	8.9
Exercise	Sedentary, %	30.5	28.1	25.3	21.3	34.4	31.3	40.0	40.0
	Regular aerobic exercise, %	10.5	8.6	14.7	13.3	7.8	6.3	8.9	4.4
Smoking	Current smoker, no. (%)	54 (28.6)	58 (29.1)	19 (26.0)	18 (24.0)	13 (21.9)	17 (26.6)	13 (35.6)	12 (33.3)
	Number of cigarettes per d (current smokers only), mean (SD)	17.7 (10.9)	14.8 (11.3)	21.0 (10.5)	18.3 (10.1)	9.9 (6.4)	8.4 (6.0)	19.3 (14.0)	16.0 (14.5)
	Former smoker, %	29.5	28.1	38.7	45.3	28.1	21.9	22.2	17.8
Alcohol	Alcohol consumption, no. (%)	116 (55.2)	109 (51.9)	56 (74.7)	54 (72.0)	24 (37.5)	24 (37.5)	22 (48.9)	18 (40.0)
	Number of drinks per mo (drinkers only), mean (SD)	17.4 (20.0)	16.2 (19.3)	20.4 (22.8)	17.2 (20.3)	14.1 (17.1)	15.3 (20.3)	11.9 (12.4)	13.3 (11.7)
	Chronic consumption, %	2.9	2.4	5.3	4.0	0.0	1.6	0.0	0.0
	Acute consumption, %	14.8	16.7	20.0	20.0	9.4	10.9	8.9	13.3
	Drink and drive, %	1.9	0.5	2.7	0.0	0.0	1.6	2.2	0.0
Obesity	Body mass index, no. reporting	200	204	74	73	61	63	40	42
	Body mass index, mean (SD)	25.9 (4.8)	26.2 (4.9)	25.6 (4.0)	25.6 (3.8)	27.9 (6.1)	27.9 (6.0)	24.7 (3.7)	26.2 (4.6)
	Obese, %	27.6	30.5	26.7	24.0	42.2	40.6	20.0	33.3
	Trying to lose weight, %	47.6	39.5	48.0	42.7	51.6	40.6	44.4	37.8
Diabetes	Diagnosed diabetes, %	6.2	5.2	5.3	4.0	10.9	10.9	4.4	2.2

Note. The total for all respondents is greater than the sum of White non-Hispanic, Black non-Hispanic, and Hispanic respondents as a result of exclusion of 26 subjects from ethnicity-specific analyses (see text for details). None of the differences between time 1 and time 2 values were significant ($P > .05$) within the whole sample or each ethnicity separately.

assess the effect on estimates of reliability of varying intervals between the two administrations.

Participation in the first interview may have prompted some respondents to modify their life-style, but this did not appear to be the case, inasmuch as there was no overall improvement in the health profile of the sample. A related issue is that if the population is homogenous with respect to the characteristic being studied, estimates of reliability become highly sensitive to small departures from perfect concordance.[8] This may explain the lower estimates of reliability for recent blood pressure check (prevalence at first administration of 96.7%), chronic alcohol consumption (prevalence of 2.9%), and driving after alcohol consumption (prevalence of 1.9%). For the race/ethnicity-specific analyses, instability due to extreme mar-

TABLE 5—Consistency of Individual Responses to Two Administrations of the Demographics Module of the Behavioral Risk Factor Surveillance System Questionnaire among 210 Respondents: Massachusetts, 1992

Demographic Characteristic	Reliability			
	All Respondents (n = 210)	White Non-Hispanic (n = 75)	Black Non-Hispanic (n = 64)	Hispanic (n = 45)
Age	1.00	1.00	1.00	.97
Sex	1.00	1.00	1.00	1.00
Education	0.80	0.89	0.76	0.62
Marital status	0.93	0.95	0.98	0.86
Employment	0.77	0.93	0.71	0.63
Income	0.75	0.82	0.65	0.67
Race/ethnicity	0.87	...	...	...

Note. Age was retained as a continuous variable. Pearson's correlation coefficient was used to estimate reliability of age. Kappa was used to estimate reliability of categorical variables. In instances in which a demographic characteristic included more than two categories, unweighted kappa was computed. See Table 1 for the number of categories used in computing kappa. All reliability coefficients were highly significant ($P < .001$). All analyses were based on the complete sample, except for age among all respondents (n = 208) and White respondents (n = 74) and education among all respondents (n = 206), Black respondents (n = 63), and Hispanic respondents (n = 43). Kappa was not computed for consistency of race/ethnicity classification within ethnicity-specific subgroups.

TABLE 6—Consistency of Individual Responses to Two Administrations of the Behavioral Risk Modules of the Behavioral Risk Factor Surveillance System Questionnaire among 210 Respondents: Massachusetts, 1992

Module	Risk Behavior/ Characteristic	Reliability			
		All Respondents (n = 210)	White Non-Hispanic (n = 75)	Black Non-Hispanic (n = 64)	Hispanic (n = 45)
Safety belt	Safety belt nonuse	0.76	0.81	0.77	0.75
Hypertension	Recent blood pressure check	0.81	0.85	−0.03	0.79
	Diagnosed hypertension	0.89	1.00	0.87	0.83
Cholesterol	Recent cholesterol check	0.77	0.83	0.83	0.57
	Hypercholesterolemia	0.86	0.88	0.86	0.73
	Cholesterol awareness	0.80	0.83	0.64	0.60
	Actual cholesterol level	0.88 (n = 42)	0.85 (n = 25)	1.00 (n = 3)	0.96 (n = 5)
Exercise	Sedentary life-style	0.57	0.59	0.50	0.63
	Regular aerobic exercise	0.45	0.61	−0.07	0.64
Smoking	Current smoker	0.83	0.90	0.79	0.85
	Number of cigarettes per d	0.73 (n = 49)	0.60 (n = 18)	0.83 (n = 13)	0.70 (n = 10)
	Former smoker	0.73	0.86	0.67	0.58
Alcohol	Any alcohol consumption	0.82	0.93	0.73	0.73
	Number of drinks per mo	0.72 (n = 103)	0.79 (n = 54)	0.57 (n = 20)	0.60 (n = 17)
	Chronic consumption	0.53	0.55	...[a]	...[a]
	Binge consumption	0.64	0.67	0.57	0.55
	Drink and drive	−0.01	...[a]	...[a]	...[a]
Obesity	Body mass index	0.90 (n = 198)	0.97 (n = 73)	0.91 (n = 61)	0.59 (n = 39)
	Obese	0.77	0.86	0.84	0.55
	Weight control	0.72	0.73	0.72	0.68
Diabetes	Diagnosed diabetes	0.82	0.85	1.00	−0.03

Note. Kappa was used to estimate reliability for dichotomous variables. Pearson's correlation coefficient was used to estimate reliability for continuous variables (actual cholesterol level, number of cigarettes per day, number of drinks per month, body mass index). Analyses were performed on data from all respondents except where specified. For continuous variables, reliability was computed only for respondents with nonzero, nonmissing values at both administrations. All kappa coefficients were highly significant ($P < .001$), except for recent blood pressure check and regular aerobic exercise among Black respondents, driving after alcohol consumption in the whole sample, and diagnosed diabetes among Hispanic respondents ($P > .05$ for each). It should be noted that the low values of kappa reflect extreme marginal prevalence of the risk behaviors rather than a large proportion of discordant cells. All correlation coefficients were significant ($P < .001$), except cholesterol level among Black ($P = .014$) and Hispanic ($P < .01$) respondents, number of cigarettes among White ($P < .01$) and Hispanic ($P = .024$) respondents, and number of drinks per month among Black ($P < .01$) and Hispanic ($P < .01$) respondents.

[a]Prevalence of zero at either time 1 or time 2, kappa not computed.

ginal prevalence was compounded by the smaller sample sizes, leading in some instances to uninterpretable negative values for the kappa statistic.

A limitation of this study is the low response rate to the initial interviews. The response rate for the main sample was consistent with that obtained in Massachusetts over the previous 3 years,[9] although it was lower than that of most other states participating in the BRFSS (C. Leutzinger, written communication, June 1992). For the minority study, the response rate to the initial interview was 39%; therefore, estimates of prevalence derived from this study may be subject to sampling bias. The response rate for the second interview was higher: 62% for the minority sample and 70% for the main sample. Thus, we are able to state with fair confidence that, among responders, consistent responses may be obtained for most parameters. We cannot assess whether reliability differs between responders and nonresponders.

In conclusion, the BRFSS questionnaire yields highly consistent group mean estimates of prevalence when administered repeatedly to the same individuals. Individual reliability is also high. This study, together with previous studies of reliability and validity,[4–6] supports the use of the BRFSS questionnaire in surveillance activities. □

Acknowledgments

This study was funded through a cooperative agreement between the Centers for Disease Control and Prevention, the Massachusetts Department of Health, and the Massachusetts Health Research Institute.

References

1. Remington PL, Smith MY, Williamson DF, Anda RF, Gentry EM, Hogelin GC. Design, characteristics and usefulness of state-based behavioral risk factor surveillance. *Public Health Rep.* 1988;103:366–375.
2. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1990. DHHS publication PHS 91-50212.
3. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960; 20:37–46.
4. Shea S, Stein AD, Lantigua R, Basch CE. Reliability of the Behavioral Risk Factor Survey in a triethnic population. *Am J Epidemiol.* 1991;133:489–500.
5. Anda RF, Dodson DL, Williamson DF, Remington PL. Health promotion data for state health departments: telephone versus in-person estimates of smoking and alcohol use. *Am J Health Promotion.* 1989;4:32–36.
6. Jackson C, Jatulis DE, Fortmann SP. The Behavioral Risk Factor Survey and the Stanford Five-City Project Survey: a comparison of cardiovascular risk behavior estimates. *Am J Public Health.* 1992;82:412–416.
7. Marcus AC, Crane LA. Telephone surveys in public health research. *Med Care.* 1986; 24:97–112.
8. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol.* 1987;126:161–169.
9. Stein AD. *Behavioral risk factors and preventive health measures—Massachusetts Behavioral Risk Factor Surveys 1986–90.* Boston, Mass: Massachusetts Dept of Public Health; 1992.

A2623-1

American Journal of Public Health

January 1994, Vol. 84, No. 1 Established 1911

ED KIANKA
HEALTH REPORTER
CONGRESSIONAL INFORMATION SVC
4520 EAST-WEST HIGHWAY
BETHESDA MD 20814
9375

Editorials

8 Cancer Prevention: Accomplishments and Prospects *P. Cole and Y. Amoateng-Adjepong*

10 The Call for Change in Breast Cancer Screening Guidelines *S. Shapiro*

12 Promoting Participation in Breast Cancer Screening *J. G. Zapka*

Articles

VARIOUS TOPICS

14 Adiposity and Stroke among Older Adults of Low Socioeconomic Status: The Chicago Stroke Study *L. DiPietro, A. M. Ostfeld, and G. L. Rosner*

20 Estimating Clinical Morbidity Due to Ischemic Heart Disease and Congestive Heart Failure: The Future Rise of Heart Failure *L. Bonneux, J. J. Barendregt, K. Meeter, G. J. Bonsel, and P. J. van der Maas*

29 Self-Reported Physical Activity in a Rural County: A New York County Health Census *C. B. Eaton, A. N. Nafziger, D. S. Strogatz, and T. A. Pearson*

33 Car Mass and Fatality Risk: Has the Relationship Changed? *L. Evans and M. C. Frick*

37 Cesarean Section and Operative Vaginal Delivery in Low-Risk Primiparous Women, Western Australia *A. W. Read, W. J. Prendiville, V. P. Dawes, and F. J. Stanley*

DISEASE PREVENTION/HEALTH EDUCATION

43 Physicians' Recommendations for Mammography: Do Tailored Messages Make a Difference? *C. S. Skinner, V. J. Strecher, and H. Hospers*

50 Mammography Screening: How Important Is Cost as a Barrier to Use? *N. Urban, G. L. Anderson, and S. Peacock*

56 Using Census and Mortality Data to Target Small Areas for Breast, Colorectal, and Cervical Cancer Screening *H. F. Andrews, J. F. Kerner, A. G. Zauber, J. Mandelblatt, J. Pittman, and E. Struening*

62 Changes in the Use of Screening Mammography: Evidence from the 1987 and 1990 National Health Interview Surveys *N. Breen and L. Kessler*

68 The Next Generation: Poor Compliance with Risk Factor Guidelines in the Children of Parents with Premature Coronary Heart Disease *N. R. Langnier, P. C. Rowe, and R. Davies*

72 "Eat For Health": A Supermarket Intervention for Nutrition and Cancer Risk Reduction *A. B. Rodgers, L. G. Kessler, B. Portnoy, A. L. Potosky, B. Patterson, J. Tenney, F. E. Thompson, S. M. Krebs-Smith, N. Breen, O. Mathews, and L. L. Kahle*

77 A Workplace Intervention for Increasing Outdoor Workers' Use of Solar Protection *A. Girgis, R. W. Sanson-Fisher, and A. Watson*

82 Racial Disparities in Reported Prenatal Care Advice from Health Care Providers *M. D. Kogan, M. Kotelchuck, G. R. Alexander, and W. E. Johnson*

89 Infant Feeding Policies in Maternity Wards and Their Effect on Breast-Feeding Success: An Analytical Overview *R. Pérez-Escamilla, E. Pollitt, B. Lonnerdal, and K. G. Dewey*

(continued page 1)

APHA
American Public Health Association

Featuring Disease Prevention/ Health Education

Calls for Abstracts and Awards Nominations

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

January 1994, Vol. 84, No. 1 ISSN: 0090-0036 Established 1911

Public Health Briefs

VARIOUS TOPICS

98 Correlates of High-Density Lipoprotein Cholesterol in Black and White Women *G. W. Heath, C. A. Macera, J. B. Croft, M. L. Mace, T. Gillette, and F. C. Wheeler*

DISEASE PREVENTION/HEALTH EDUCATION

101 Cancer Screening among African-American Women: Their Use of Tests and Social Support *S. H. Kang, J. R. Bloom, and P. S. Romano*

104 Promoting Mammography Use through Progressive Interventions: Is It Effective? *E. S. King, B. K. Rimer, J. Seay, A. Balshem, and P. F. Engstrom*

107 The Use of Mammography Vans by Low-Income Women: The Accuracy of Self-Reports *S. Etzi, D. S. Lane, and R. Grimson*

110 Screening for Childhood Lead Poisoning: A Cost-Minimization Analysis *D. E. Glotzer, H. Bauchner, K. A. Freedberg, and S. Palfrey*

113 Who Participates in Cardiovascular Disease Risk Factor Screenings? Experience with a Religious Organization–Based Program *B. L. Wells, C. C. Brown, J. W. Horm, R. A. Carleton, and T. M. Lasater*

116 Is Cardiopulmonary Resuscitation Training Deleterious for Family Members of Cardiac Patients? *K. Dracup, D. K. Moser, P. M. Guzy, S. E. Taylor, and C. Marsden*

119 Educating Patients with Limited Literacy Skills: The Effectiveness of Printed and Videotaped Materials about Colon Cancer *C. D. Meade, W. P. McKinney, and G. P. Barnas*

Other Journal Departments

2 APHA Masthead

6 What *AJPH* Authors Should Know

7 Advertisers' Index

8 *AJPH* Masthead

122 Letters to the Editor

125 Erratum

126 Book Corner

130 Association News: Call for Abstracts—APHA 122nd Annual Meeting in Washington, DC

143 Association News: Nominations Invited for 1994 APHA Awards

153 Job Opportunities

APHA

American Public Health Association

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 Fifteenth Street, NW, Washington, DC 20005; (202) 789-5600. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*

Helen Rodriguez-Trias, MD, *Immediate Past President*

Caswell A. Evans, Jr, DDS, MPH, *President-Elect*

Leonard M. Schuman, MD, *Vice-President (USA)*

John E. F. Hastings MD, MPH, *Vice-President (Canada)*

Oscar Feo, MD, MSc, *Vice-President (Latin America)*

Jay H. Glasser, PhD, *Treasurer*

Allan G. Rosenfield, MD, *Chair, Executive Board*

Shirley F. Randolph, MSPH, *Speaker, Governing Council*

Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*

Carol Easley Allan, PhD, RN (1997)

Caswell A. Evans, Jr, DDS, MPH, *President-Elect*

Patricia P. Evans, MPH (1994)

Eugene Feingold, PhD, JD, *President*

Jay H. Glasser, PhD, *Treasurer*

Alan R. Hinman, MD, MPH (1995)

Margo D. Kerrigan, MPH (1995)

Joanne Lukomnik, MD (1996), *Vice-Chair*

Steven Uranga McKane, DMD, MPH (1995)

Carlos W. Molina, EdD (1997)

Linda Rae Murray, MD, MPH (1996)

Hila Richardson, DrPH, MPH (1996)

Helen Rodriguez-Trias, MD, *Immediate Past President*

Reed V. Tuckson, MD (1994)

Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*

Frances Kissling, *Chair, Action Board*

George E. Pickett, MD, MPH, *Speaker, Program Development Board*

Shirley F. Randolph, MSPH, *Speaker, Governing Council*

American Journal of Public Health

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, Inc, 1015 Fifteenth St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 Fifteenth St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 Fifteenth St, NW, Washington, DC 20005. AJPH allotted $40; TNH allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 Fifteenth St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, AJPH Reprints, APHA, 1015 Fifteenth St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The American Journal of Public Health is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

Adiposity and Stroke among Older Adults of Low Socioeconomic Status: The Chicago Stroke Study

Loretta DiPietro, PhD, MPH, Adrian M. Ostfeld, MD, and Gary L. Rosner, ScD

ABSTRACT

Objectives. **The purpose of this study was to test the hypothesis that overall and truncal adiposity increase the risk of stroke independent of their association with cardiovascular disease risk factors and other preexisting illnesses.**

Methods. **Analyses were conducted of longitudinal data from a poor, biracial cohort of noninstitutionalized adults 65 to 74 years of age who participated in the Chicago Stroke Study from 1965 to 1970.**

Results. **Ponderal index (cm/kg$^{1/3}$) and chest skinfold were significantly associated with systolic and diastolic blood pressure, serum cholesterol and triglycerides, plasma glucose, and smoking. Ponderal index was also associated with diabetes and risk of stroke. After potential confounders were controlled, the following variables showed significant independent associations with risk of stroke: Black race, female gender, and age 70+; hypertensive heart disease; and diabetes. Neither adiposity variable was associated with risk of stroke in the presence of these powerful predictors.**

Conclusions. **Control of hypertension and diabetes continues to be important among older adults. Since excess adiposity seems to influence risk of stroke through its association with these disorders and other cardiovascular disease risk factors, control of weight and fat remains an important concern as well. (*Am J Public Health.* 1994;84:14–19)**

Introduction

Epidemiologic studies of the relationship between overall obesity and cardiovascular disease have yielded inconclusive results.[1–6] Recent studies that have focused specifically on fat distribution, however, have demonstrated an increased risk of cardiovascular disease or mortality with increasing levels of abdominal adiposity.[7–11]

Adiposity gained in later adulthood tends to accumulate on the trunk and upper body rather than the periphery,[12] particularly among postmenopausal women.[12] An android type of fat patterning, characterized by an excess of upper-body fat relative to lower-body fat[13] and composed of hypertrophied fat cells, is both diabetogenic and atherogenic.[13–20] Since diabetes, hypertension, and altered plasma lipid profiles are primary risk factors for cardiovascular disease, excess truncal fat may serve as a more basic risk factor in its etiology. Whether overall adiposity or truncal adiposity exerts an independent effect on risk of cardiovascular disease remains unclear.

Most of the studies on adiposity and cardiovascular disease have been limited to middle-aged populations at highest risk of coronary heart disease. Much less is known about how adiposity influences the risk of cerebrovascular disease among older adult populations. This is especially true for older Black women, among whom the prevalence of obesity, as well as hypertension and diabetes, tends to be disproportionally high relative to older men and women of other races.[21,22]

In order to examine the multivariable association between a measure of overall adiposity (the ponderal index), a measure of truncal adiposity (the chest skinfold), and the risk of stroke among a poor, biracial, older adult population, we analyzed data from a cohort of individuals who participated in the Chicago Stroke Study.[23] Earlier work by Ostfeld et al.[24] demonstrated a simple relationship between the ponderal index and stroke in this Chicago cohort. The primary hypothesis tested in the current analysis was that both overall adiposity and truncal adiposity increase the risk of stroke among older adults, independent of their association with cardiovascular disease risk factors and preexisting illness other than cardiovascular disease.

Methods

The Chicago Stroke Study was conducted from 1965 to 1970 with three probability samples of noninstitutionalized adults who were 65 to 74 years of age, lived in Cook County, Illinois, and were receiving old age assistance (public assistance to those with virtually no income and with possessions worth less than $400). Eligible subjects were classified into four strata by race (White/Black) and sex. Sampling fractions were used to secure equal numbers of each race–sex stratum. The sampling methods have been described in detail elsewhere.[23]

Loretta DiPietro and Adrian M. Ostfeld are with the School of Medicine, Department of Epidemiology and Public Health, Yale University, New Haven, Conn. At the time this article was written, Gary L. Rosner was also with the School of Medicine, Department of Epidemiology and Public Health, Yale University. He is now with the Duke University Medical Center, Department of Cancer and Biostatistics, Durham, NC.

Requests for reprints should be sent to Loretta DiPietro, PhD, MPH, The John B. Pierce Laboratory, 290 Congress Ave, New Haven, CT 06519.

This paper was accepted May 11, 1993.

Of the 4759 individuals in the original target population, 9% could not be located, 19% refused to participate, and 6% died in the interval between the time of sampling and the initial medical examination. Therefore, 3141 (66%) subjects underwent the initial examination. Baseline examinations were conducted between September 1965 and February 1967. Since the primary purpose of the study was to determine the factors affecting the risk of stroke, those individuals with a prior history of cerebrovascular injury (n = 368) were excluded from the analysis. Therefore, the study population comprised 2773 persons.

The period of follow-up was between September 1965 and March 1970; the median follow-up period for the study cohort was 32.1 months. During the follow-up period, 21% of the subjects died, 3% moved from the Chicago area, and 0.7% could not be traced.

The initial history and medical examinations were virtually identical to those of the Framingham study, except for inclusion of items germane to examination of the nervous system and determination of the risk factors for stroke. Additional diagnostic procedures included a 12-lead electrocardiogram, retinoscopy, blood lipid levels, serum glucose levels obtained 1 hour after oral intake of 50 g of glucose, systolic and diastolic blood pressure, forced expiratory volume for 1 second, height, weight, and chest skinfold thickness measured over the lateral surface of the 10th rib.[24]

Categories of preexisting clinical cardiovascular disease upon initial examination included myocardial infarction, angina pectoris, congestive heart failure, peripheral arterial disease, hypertensive heart disease, and stroke. Subjects were divided into these diagnostic categories based on criteria previously outlined by Ostfeld et al.[24] All persons with clinical heart disease, transient ischemic attacks, peripheral atherosclerosis, intermittent claudication, definite electrocardiogram abnormalities, history of diabetes mellitus, blood urea nitrogen of 30 mg/dl or more, or health status rated as gravely ill at baseline were classified as Population A. Population B consisted of subjects without any of these conditions. This stratification was done to determine whether the various risk factors for stroke operated independently of cardiovascular or closely related diseases.

Definition of Stroke

Individuals free of a prior stroke upon the initial medical examination were followed until a stroke or death occurred or until March 1970, when the follow-up period ended. In order to identify those subjects who developed new cerebrovascular injuries, all subjects were visited biannually by field workers trained to observe signs of strokes and transient ischemic attacks. A trained research supervisor evaluated these data against strict criteria for selecting individuals for follow-up examination by a board-certified neurologist (88% of the subjects who experienced signs and symptoms of stroke and who survived were examined by the neurologist). As long as the information was negative for the possibility of a stroke, the subject was reevaluated at the next biannual exam. The Cook County Hospital provided another source of morbidity and mortality data, since approximately 88% of hospitalizations among the study population occurred there. Admissions to this hospital were screened daily to determine whether subjects were hospitalized, the diagnosis, and, if relevant, the cause of death. The same morbidity and mortality data were collected from other hospitals in which cohort members were admitted, but this information was obtained at wider intervals.

All diagnostic decisions were made by a neurologist according to criteria established at the onset of the study. For living subjects, decisions were based on examinations performed by a neurologist or senior resident. Diagnostic decisions for deceased individuals were based on medical and hospital records, death certificates, or autopsy reports. Differentiations between brain infarct and hemorrhage were made whenever sufficient data were available.

Adiposity

Overall baseline adiposity was assessed by the ponderal index ($cm/kg^{1/3}$) in order to remain consistent with previously published reports from these data.[24] The ponderal index is inversely associated with adiposity and correlated almost perfectly with both body mass index (kg/m^2) ($r = -.96$) and relative weight (percentage of ideal weight) ($r = -.96$) among our study population. For a man of average height, a ponderal index of 41.8 $cm/kg^{1/3}$ corresponds to a body mass index of 23.8 (percentage of ideal weight = 100); a ponderal index of 39.6 $cm/kg^{1/3}$ corresponds to a body mass index of 27.8 (percentage of ideal weight = 120), which is a cutpoint often used to define obesity.[25] Truncal adiposity was assessed at baseline by a chest skinfold measured over the lateral surface of the 10th rib with Best calipers[26] and recorded in millimeters. Accurate skinfolds could not be measured for 200 subjects (7%) because the fat layer surrounding the 10th rib was larger than what the calipers could measure. Accordingly, the skinfolds of these subjects were recorded as 88 mm (the highest reading on the calipers), and the chest skinfold was categorized on the basis of quartiles of its distribution (<7 mm, 7 through 11.9 mm, 12 through 18.9 mm, or ≥19 mm).

Analysis

Univariate statistics were first generated on all study variables. A stroke was the outcome of interest and was first considered as a dichotomous variable. Pearson product–moment correlations, analysis of variance, the maximum likelihood ratio chi-square, and logistic regression[27] tested the simple associations among risk factors and between risk factors and the outcome.

We next considered time until the event of a stroke, measured in years from the start of the study, as the outcome variable. Cox proportional hazards modeling[28] was used to test the associations between the study variables of interest and rate of stroke while simultaneously controlling for important covariates and possible confounders, including preexisting disease states (Appendix). Two separate models were considered in the multivariable analysis, one containing the ponderal index measure and the other containing the chest skinfold measure.

In order to account for the sampling strata and to assess race-, sex-, and age-specific differences in risk of stroke, the dichotomous variables race, sex, and age (<70/≥70) were combined to form an eight-level interaction variable. This race–sex–age variable was included in all simple and multivariable analyses as a categorical indicator variable. A variable defining high blood pressure (yes/no) was used in place of both systolic and diastolic blood pressure in the multivariable Cox modeling procedures to test their threshold effect in predicting stroke. High blood pressure was scored 0 if systolic blood pressure was less than 160 mm Hg and diastolic blood pressure was less than 95 mm Hg, and it was scored 1 if otherwise; these cutpoints conform to the standard definition of high blood pressure at the time these data were collected. Since less than 10% of our population was on antihypertensive medication (and, among the users, adherence was generally low), use

TABLE 1—Mean Ponderal Index and Body Mass Index and Percentage Distribution of Chest Skinfold Quartiles by Race, Sex, and Age: Chicago Stroke Study

	Men		Women	
	White	Black	White	Black
Age <70 y				
Ponderal index,[a] mean (SE)	41.4 (0.16)	41.9 (0.14)	39.1 (0.21)	38.8 (0.18)
Body mass index,[a] mean (SE)	24.2 (0.29)	23.6 (0.23)	26.7 (0.41)	28.2 (0.38)
Chest skinfold,[b] % distribution				
<7 mm	43.7	51.9	19.4	24.6
7–11.9 mm	23.2	21.8	18.5	10.6
12–18.9 mm	21.7	18.1	26.7	27.6
≥19 mm	11.4	8.2	35.3	37.0
Age ≥70 y				
Ponderal index,[a] mean (SE)	41.6 (0.16)	41.7 (0.15)	38.8 (0.16)	39.3 (0.15)
Body mass index,[a] mean (SE)	23.6 (0.26)	24.1 (0.25)	27.1 (0.31)	27.1 (0.30)
Chest skinfold,[b] % distribution				
<7 mm	48.7	49.2	20.2	22.5
7–11.9 mm	20.9	18.6	16.8	18.4
12–18.9 mm	18.7	18.6	32.8	28.0
≥19 mm	11.7	13.6	30.1	31.1

[a]Overall race–sex–age effect ($P < .0001$) based on analysis of variance; pairwise comparisons based on studentized maximum modulus (GT^2) test.
[b]Overall race–sex–age effect ($P < .0001$) based on likelihood ratio chi-square (G^2).

TABLE 2—Associations between Ponderal Index, Quartiles of Chest Skinfold, and Mean Levels of Selected Cardiovascular Disease Risk Factors: Chicago Stroke Study

	Ponderal Index		Chest Skinfold	
	r^a	P	R^{2b}	P
Systolic blood pressure, mm Hg	−.17	.0001	.02	.0001
Diastolic blood pressure, mm Hg	−.16	.0001	.02	.0001
Serum cholesterol, mg/dl[c]	−.22	.0001	.04	.0001
Triglycerides, mg/dl[c]	−.20	.0001	.06	.0001
Plasma glucose, mg/dl	−.11	.0001	.01	.0001
Blood urea nitrogen, mg/dl	−.06	.003	.00	.58
FEV, % FVC	−.32	.0001	.05	.0001

Note. FEV = forced expiratory volume for 1 second expressed as a percentage of forced vital capacity (FVC).
[a]Pearson product–moment correlation coefficient.
[b]Square of the coefficient of correlation based on analysis of variance.
[c]n = 2433.

of such medication was not considered in the definition of high blood pressure.

Odds ratios (ORs) based on the logistic regression estimates and rate ratios based on the Cox regression estimates, along with 95% confidence intervals (CIs), described the magnitude of effect for each level of the covariates compared with the reference category. All statistical procedures used the SAS (SAS Institute, Cary, NC) and BMDP (University of California, Berkeley) computer packages.

Results

Demographic and risk factor characteristics of the study population have been published previously by Ostfeld et al.[23]

Stroke

One hundred ninety-nine strokes occurred during the follow-up period. For 136 (68%) of these strokes, there was a clear history with confirming signs (designated as class I infarcts); 7 (3%) had a clear history without signs, 16 (8%) had a vague history with signs, 11 (6%) showed signs only, and 29 (15%) were determined from a death certificate. In order to be as inclusive as possible, all 199 strokes were included in the analysis. Most of these 199 strokes (80%) were nonembolic infarcts; 11% were hemorrhages, 2% were emboli, and 7% were undetermined.

Adiposity

In general, women had the lowest mean ponderal index (highest level of overall adiposity) in the study population (Table 1), and Black women had the highest prevalence of ponderal indices less than 38.1 (the lowest quartile). Differences in ponderal index between White and Black women were not statistically significant in either age category. Under age 70, women had a significantly lower ponderal index than did men of both races. After age 70, however, only White women showed significant differences in ponderal index from their male counterparts. Ponderal index did not change substantially with age. Also, women had more fat surrounding the chest wall than did men (Table 1). Black women had the highest proportion of skinfold measures greater than 19 mm, which represents the highest quartile of the distribution.

Adiposity and Other Risk Factors for Stroke

Both ponderal index and chest skinfold were significantly associated with systolic blood pressure, diastolic blood pressure, serum cholesterol, serum triglycerides, and plasma glucose (Table 2), although these correlations were of a rather modest magnitude. Ponderal index ($P < .0001$) and chest skinfold ($P < .0001$) showed significant inverse associations with current smoking in the presence of race, sex, and age. Ponderal index was also significantly associated with a history of diabetes while the race–sex–age variable was controlled for (OR = 0.95, 95% CI = 0.93, 0.98), whereas chest skinfold was not.

Risk Factors for Stroke

Overall, we found that the risk of stroke varied substantially by race, sex, and age in the simple analysis. Among subjects less than 70 years of age, Black men and women had more than twice the risk of stroke than White men (the referent category); among those 70 years of age or older, Black women had the highest risk of stroke (more than 3.5 times that of younger White men) (Table 3). Also, among subjects 70 years of age or older, the risk of stroke was twice as high for White subjects of both sexes as for their younger counterparts.

The following variables were significantly associated with the risk of stroke in the presence of the race–sex–age stratification variable alone: population code (B/A) (OR = 0.49, 95% CI = 0.35, 0.68), diabetes (OR = 1.81, 95% CI = 1.31, 2.53), hypertensive heart disease (OR = 2.52, 95% CI = 1.85, 3.41), left ventricular hypertrophy (OR = 1.41, 95% CI = 1.01, 1.99), hypertensive retinopathy (grade 1:

OR = 1.97, 95% CI = 1.33, 2.90; grades 2 through 4: OR = 2.51, 95% CI = 1.74, 3.65), and blood urea nitrogen (OR = 1.02, 95% CI = 1.01, 1.04). Ponderal index was also related to the risk of stroke, although the association was of marginal statistical significance (OR = 0.95, 95% CI = 0.90, 1.00). Chest skinfold, systolic and diastolic blood pressure, serum cholesterol, serum triglycerides, plasma glucose, peripheral vascular disease, forced expiratory volume for 1 second, and smoking were not associated with risk of stroke while adjusting for age, sex, and race.

After controlling for all covariates simultaneously and substituting the variable high blood pressure for both systolic and diastolic blood pressure, we found that the study variables ponderal index and chest skinfold showed no independent effect on the rate of stroke (Table 4). Black women 70 years of age or older continued to experience a substantially higher rate of stroke than did White men less than 70 years old. The rate of stroke among individuals with hypertensive heart disease was more than twice that of healthier individuals; those with diabetes experienced nearly a 70% higher rate. Subjects who met our criteria for high blood pressure had a 30% higher rate of stroke than did subjects without high blood pressure, although this association did not reach statistical significance.

TABLE 3—Risk of Stroke, by Race, Sex, and Age and Based on Simple Logistic Regression Modeling: Chicago Stroke Study

	n	Frequency of Stroke	Crude Risk	Odds Ratio	95% Confidence Interval
Age <70 y					
White men	272	9	.033	1.00	...
Black men	376	25	.066	2.08	0.95, 4.53
White women	221	9	.041	1.24	0.48, 3.24
Black women	311	20	.064	2.01	0.90, 4.51
Age ≥70 y					
White men	316	22	.070	2.19	0.99, 4.85
Black men	360	28	.077	2.46	1.14, 5.31
White women	405	31	.076	2.42	1.13, 5.24
Black women	512	55	.107	3.52	1.71, 7.23

TABLE 4—Associations between Ponderal Index, Chest Skinfold, Selected Cardiovascular Disease Risk Factors, and Stroke Based on Cox Proportional Hazards Models: Chicago Stroke Study

	Model 1		Model 2	
	Relative Risk	95% Confidence Interval	Relative Risk	95% Confidence Interval
Ponderal index	0.96	0.84, 1.08		
Chest skinfold				
<7 mm			1.00	...
7–11.9 mm			0.88	0.58, 1.33
12–18.9 mm			0.99	0.69, 1.44
≥19 mm			0.83	0.56, 1.24
Black women aged ≥70 y	2.38	1.16, 4.89	2.39	1.16, 4.91
Hypertensive heart disease	2.18	1.58, 2.99	2.17	1.58, 2.98
Diabetes	1.66	1.22, 2.27	1.66	1.22, 2.27
High blood pressure	1.27	0.94, 1.73	1.28	0.94, 1.74

Note. Two separate models were considered, one with the ponderal index and the other with chest skinfold.

Discussion

Several longitudinal studies have reported overall adiposity and, specifically, adiposity distributed around the trunk to be significant and independent risk factors for cardiovascular disease.[7–10] Whether this excess risk contributed by adiposity continues into older adulthood is not clear.

The main hypothesis of this study was not substantiated. Overall adiposity and truncal adiposity did not serve as independent predictors of stroke in our study population. The ponderal index showed a marginally significant association with stroke incidence in the simple analysis but failed to retain statistical significance in the presence of the other covariates. Black women 70 years of age or older and individuals with preexisting hypertensive heart disease and diabetes continued to demonstrate an excess rate of stroke in the multivariable analysis. These findings suggest that adiposity exerted its effects on stroke through its association with stroke-related disorders and risk factors among our older population rather than through a direct, independent association.

Three longitudinal studies[24,29,30] found measures of overall adiposity to be associated with an increased risk of stroke among older populations. Whether or not these associations were independent of such factors as hypertension or diabetes is unclear. There is some evidence from the Framingham study[31] that relative weight is associated with increased risk of stroke among women only, independent of blood pressure. Among subjects in the Evans County study,[22] body mass index was an independent predictor of stroke mortality only among White women of lower socioeconomic status.

The Framingham study[10] also found a positive relationship between its indicator of truncal adiposity (subscapular skinfold thickness) and 28-year incidence of coronary heart disease that was independent of body mass index. Other studies have found subscapular skinfold thickness, or a combination of skinfold measures, to be associated with risk factors for cardiovascular disease such as blood pressure, serum lipids, and serum glucose.[32,33]

The causal mechanisms for the relationship between truncal adiposity and cardiovascular disease are not fully understood. An android type of fat patterning, characterized by an excess of upper-body fat relative to lower-body fat and composed of hypertrophied fat cells,[13] is related to various hemodynamic and metabolic complications, such as hyperlipidemia, hyperinsulinemia, glucose intolerance, and sodium retention.[14–20] The consequences of these various aberrations for conditions such as diabetes, hypertension, and subsequent cardiovascular disease seem clear, although evidence for a direct or indirect causal pathway has not been confirmed.

Adiposity may be important to disease conditions such as hypertensive heart disease or diabetes but may have explained nothing over and above these powerful predictors of stroke among our older study population. Moreover, by controlling for the variables through which obesity mediates its effect on car-

diovascular disease, we may have underestimated the effect of obesity.

These results from the Chicago Stroke Study data also suggest that the excess cardiovascular risk associated with adiposity in middle age loses its impact in older ages.[34] Several studies have identified age as another important modifying variable of stroke incidence.[35,36] While adiposity was associated with traditional cardiovascular disease risk factors, such as blood pressure, serum lipid levels, glucose levels, and smoking, these correlations were weaker than those found among younger populations.[37,38] Moreover, these traditional cardiovascular disease risk factors were not significantly related to the risk of stroke among our older population in the presence of age, race, and sex; however, their disease sequelae (hypertensive heart disease and diabetes) were. The finding that neither systolic nor diastolic blood pressure was associated with stroke in the presence of age, sex, and race alone suggests that target organ damage may be necessary to elevate risk of stroke among a cohort in which mechanisms of selective survival are operating. More knowledge is needed about the natural history of blood pressure, hypertension, and cardiovascular disease with aging.

Criticisms of other studies finding negative results between adiposity and cardiovascular disease among older adult populations have focused on the failure of these studies to control for such suspected confounders as smoking, preexisting disease, antihypertensive medication, and early mortality.[39] Our analysis took these factors into account and therefore can rule out their attenuating effect on the adiposity and stroke relationship in our study population.

Conclusion

Control of hypertension and diabetes continues to be important among older adults. Since excess adiposity seems to exert its influence on the risk of stroke through its association with these preexisting disorders, control of excess weight and fat remains an important concern among this age group as well. Prevention programs aimed at weight and hypertension control (through diet and increasing physical activity) would seem most efficacious in lowering cardiovascular and cerebrovascular disease risk and need to be made accessible to all segments of our population. □

Acknowledgments

We gratefully acknowledge the contributions of Robert F. Anda, MD, and Albert J. Stunkard, MD, in their review of the manuscript.

References

1. Keys A. Overweight, obesity, coronary heart disease, and mortality. *Nutr Rev.* 1980;38:297–307.
2. Larsson B, Bjorntorp P, Tibblin G. The health consequences of moderate obesity. *Int J Obes.* 1981;5:97.
3. Barrett-Connor EL. Obesity, atherosclerosis, and coronary artery disease. *Ann Intern Med.* 1985;103:1010–1019.
4. Hubert HB. The importance of obesity in the development of coronary risk factors and disease: the epidemiologic evidence. *Annu Rev Public Health.* 1986;7:493–502.
5. Lapidus L, Bengtsson C. Regional obesity as a health hazard in women—a prospective study. *Acta Med Scand.* 1988; 723(suppl):53–59.
6. Manson JE, Colditz G, Stampfer MJ, et al. A prospective study of obesity and risk of coronary heart disease in women. *N Engl J Med.* 1990;322:882–889.
7. Larsson B, Svardsudd K, Welin L, et al. Abdominal adipose tissue distribution, obesity, and risk of cardiovascular disease and death: a 13-year follow-up of participants in the study of men born in 1913. *BMJ.* 1984;288:1401–1404.
8. Lapidus L, Bengtsson C, Larsson B, et al. Distribution of adipose tissue and risk of cardiovascular disease and death: a 12-year follow-up of participants in the population of women in Gothenburg, Sweden. *BMJ.* 1984;289:1257–1261.
9. Donahue RP, Abbott RD, Bloom E, et al. Central obesity and coronary heart disease in men. *Lancet.* 1987;i:821–824.
10. Higgins M, Kannel W, Garrison R, Pinsky J, Stokes J. Hazards of obesity—the Framingham experience. *Acta Med Scand.* 1988;723(suppl):23–36.
11. Stern MP, Patterson JK, Mitchell BD, Haffner SM, Hazuda HP. Overweight and mortality in Mexican Americans. *Int J Obes.* 1990;14:623–629.
12. Roche AF, Abdel-Malek AK, Mukherjee D. New approaches to clinical assessment of adipose tissue. In: *Body Composition Assessments in Youth and Adults.* Columbus, Ohio: Ross Laboratories; 1985.
13. Vague J. The degree of masculine differentiation of obesities: a factor determining predisposition to diabetes, atherosclerosis, gout, and uric calculus disease. *Am J Clin Nutr.* 1956;4:20–34.
14. Messerli FH, Christie B, DeCarvalho JGR, et al. Obesity and essential hypertension: hemodynamics, intravascular volume, sodium excretion and plasma renin activity. *Arch Intern Med.* 1981;141:81–85.
15. Vague J, Bjorntorp P, Guy-Grand B, et al. *Metabolic Complications of Human Obesity.* Amsterdam, the Netherlands: Elsevier Science; 1985:287.
16. Kissebah AH, Peiris AN. Mechanisms associating body fat distribution to glucose intolerance and diabetes mellitus: window with a view. *Acta Med Scand.* 1988; 723(suppl):79–89.
17. Despres J-P, Moorjani S, Lupien PL, et al. Regional distribution of body fat, plasma lipoproteins, and cardiovascular disease. *Arteriosclerosis.* 1990;10:497–511.
18. Bjorntorp P. Possible mechanisms relating fat distribution and metabolism. In: Bouchard C, Johnston FE, eds. *Fat Distribution during Growth and Later Health Outcomes.* New York, NY: Alan R Liss; 1988.
19. Kissebah AH, Vydellingum N, Murray R, et al. Relations of body fat distribution to metabolic complications of obesity. *J Clin Endocrinol Metab.* 1982;54:254–260.
20. Stern MP, Haffner SM. Body fat distribution and hyperinsulinemia as risk factors for diabetes and cardiovascular disease. *Arteriosclerosis.* 1986;6:123–130.
21. Established Populations for Epidemiologic Studies of the Elderly. *Resource Data Book.* Bethesda, Md: National Institute on Aging; 1987. DHHS publication NIH 86-2443.
22. Johnson JL, Heineman EF, Heiss G, et al. Cardiovascular disease risk factors and mortality among Black women and White women, age 40–64 years in Evans County, Georgia. *Am J Epidemiol.* 1986;123:209–220.
23. Ostfeld AM, Shekelle RB, Tufo HM, et al. Cardiovascular and cerebrovascular disease in an elderly, poor urban population. *Am J Public Health.* 1971;61:19–29.
24. Ostfeld AM, Shekelle RB, Klawans H, Tufo HM. Epidemiology of stroke in an elderly welfare population. *Am J Public Health.* 1974;64:450–458.
25. Burton BT, Forster WR, Hirsch J, et al. Health implications of obesity: an NIH consensus development conference. *Int J Obes.* 1985;9:155–169.
26. Best WR. An improved caliper for measurement of skinfold thickness. *USAN-RNL Rep.* August 1953; no. 13.
27. Bishop Y, Feinberg S, Holland P. *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass: MIT Press; 1980.
28. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data.* New York, NY: John Wiley & Sons Inc; 1980.
29. Paffenbarger RS, Williams V. Chronic disease in former college students. V. Early precursors of fatal stroke. *Am J Public Health.* 1967;57:1290–1299.
30. Comstock GW, Kendrick MA, Livesay VT. Subcutaneous fatness and mortality. *Am J Epidemiol.* 1966;83:548–563.
31. Hubert HB, Feinleib M, McNamara DM, et al. Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation.* 1983;67: 968–977.
32. Blair D, Habicht J-P, Sims EAH, et al. Evidence for an increased risk of hypertension with centrally located body fat and the effect of race and sex on this risk. *Am J Epidemiol.* 1984;119:526–540.
33. Reichley KB, Mueller WH, Hanis CL, et al. Centralized obesity and cardiovascular disease risk factors in Mexican Americans. *Am J Epidemiol.* 1987;125:373–386.
34. Larsson B. Obesity, fat distribution and cardiovascular disease. *Int J Obes.* 1991; 15:53–57.
35. Pooling Project Research Group. Rela-

tionship of blood pressure, serum cholesterol, smoking habits, relative weight, and ECG abnormalities to incidence of major coronary events: final report of the pooling project. *J Chronic Dis.* 1978;31: 201–306.
36. Ostfeld AM. A review of stroke epidemiology. *Epidemiol Rev.* 1980;2:136–152.
37. Kannel WB, Gordon T, Castelli WP. Obesity, lipids, and glucose tolerance. The Framingham Study. *Am J Clin Nutr.* 1979; 32:1238–1245.
38. Gillum RF, Taylor HL, Brozek J, et al. Indices of obesity and blood pressure in young men followed 32 years. *J Chronic Dis.* 1982;35:211–219.
39. Manson J[illegible]stampfer MJ, Hennekens CH. Body bu[illegible] and longevity: a reassessment. *JAMA.* 1987;257:353–358.
40. Fisher CM, Mohr JP, Adams RD. *Harrison's Principles of Internal Medicine.* New York, NY: McGraw-Hill; 1974.

APPENDIX—Covariates of Interest in the Chicago Stroke Study

Systolic blood pressure (mm Hg)
Diastolic blood pressure (mm Hg)
Serum cholesterol (mg/dl)
Nonfasting serum triglycerides (mg/dl)
Plasma glucose (mg/dl)
Blood urea nitrogen (mg/dl)
FEV (% FVC)
Population code (A, B)
Hypertensive heart disease (no, yes)
Peripheral vascular disease (no, yes)
Left ventricular hypertrophy (no, yes)
Hypertensive retinopathy (none, grade 1, grades 2–4)[a]
Diabetes (no, yes)
Current cigarette smoking (none, ≤0.5 pack/day, >0.5 pack/day)

Note. FEV = forced expiratory volume for 1 second, expressed as a percentage of forced vital capacity (FVC).
[a]Based on Keith-Wagener grades.[40]

Planning Themes in the Journal: A Call for Papers

We are engaged in a small experiment. For three upcoming issues, we have announced the featured theme ahead of time. To ensure the submission of a sufficient number of papers, we invite potential authors among our readers to submit papers relevant to the chosen topics by the dates announced.

In recent volumes, Journal editors have successfully assembled themes out of the material on hand. Reader response has been positive. Theme issues announced in advance will be more coherent than those created from random submissions if they attract the number of authors we hope for. They will also enable us to indicate the Journal's interest in neglected areas that authors may not see as high on our agenda now.

Our first call for papers, announcing three topics, was published in the November 1992 issue of the Journal. Papers submitted for the first theme, "Children: Societal and Individual Violence, Injury, and Abuse" (deadline was April 1, 1993), have been reviewed; those revised and accepted will be published in the April 1994 issue. Papers submitted for the second theme, "Age and Aging: Epidemiology, Health Care, and General Public Health" (deadline was August 1, 1993), have been reviewed; those revised and accepted will be published in mid-1994. (Please note that submissions on these themes are welcome at any time.)

The third theme and its deadline for submission is as follows:

Primary Care and Public Health
Submissions due February 1, 1994

Before sending us your manuscript, please consult "What *AJPH* Authors Should Know," published in each issue of the Journal and available from the office, for standards of submission and address.

Estimating Clinical Morbidity Due to Ischemic Heart Disease and Congestive Heart Failure: The Future Rise of Heart Failure

ABSTRACT

Objectives. Many developed countries have seen declining mortality rates for heart disease, together with an alleged decline in incidence and a seemingly paradoxical increase in health care demands. This paper presents a model for forecasting the plausible evolution of heart disease morbidity.

Methods. The simulation model combines data from different sources. It generates acute coronary event and mortality rates from published data on incidences, recurrences, and lethalities of different heart disease conditions and interventions. Forecasts are based on plausible scenarios for declining incidence and increasing survival.

Results. Mortality is postponed more than incidence. Prevalence rates of morbidity will decrease among the young and middle-aged but increase among the elderly. As the milder disease states act as risk factors for the more severe states, effects will culminate in the most severe disease states with a disproportionate increase in older people.

Conclusions. Increasing health care needs in the face of declining mortality rates are no contradiction, but reflect a tradeoff of mortality for morbidity. The aging of the population will accentuate this morbidity increase. (*Am J Public Health.* 1994;84: 20–28)

Luc Bonneux, MD, MSc, Jan J. Barendregt, MA, Karin Meeter, MD, PhD, Gouke J. Bonsel, MD, PhD, and Paul J. van der Maas, MD, PhD

Introduction

Heart disease, comprising ischemic heart disease and congestive heart failure, is not only the leading cause of death and lost life expectancy in most Western countries but also one of the most important causes of morbidity and health care costs.[1] After a period of increasing mortality during the 1950s and 1960s, Western countries have witnessed a dramatic decline of ischemic heart disease mortality (age-adjusted), with the United States taking the lead.[2] Though the dynamics are not completely understood, time trends in known risk factors for ischemic heart disease (hypertension, cholesterol, smoking) in Western countries indicate a clear decline.[3–5] Therefore it is reasonable to attribute at least part of this decrease in mortality to a decrease in incidence.

The observed sharp decline in mortality, together with the probable decline in incidence, leads one to expect a concomitant decrease in health care utilization. In reality, however, health care providers all over the Western world are facing increases in demand. After the introduction of percutaneous transluminal coronary angioplasties, utilization of these procedures increased almost exponentially, while the growth in utilization of coronary artery bypass graft surgery merely slowed.[6] Viewed superficially, the observed increases in demand are incongruent with the supposed decrease in need.

But is need truly decreasing? Basic epidemiologic theory teaches us that if mortality slows faster than incidence, prevalence will increase. Although the Framingham Heart Study has shown a small decline in incidence, a strong increase in prevalence has been observed, indicative of a survival gain.[3] This improved prognosis has been caused partly by improved risk factor distributions, but medical therapy has also contributed.[7,8] Recently, beta-blocking agents, platelet aggregation inhibitors, coronary artery bypass graft surgery, and angiotensine converting enzyme inhibitors have been applied in chronic therapy and thrombolytic therapy has been applied in emergencies. All have been proven effective in randomized controlled trials meeting high standards.[9–11] If trends in risk factors and survival remain favorable, the further decrease in incidence and mortality of ischemic heart disease is likely to be accompanied by a continuing increase in prevalence. In Western countries, the increased prevalence will be accentuated by the aging of the baby boom generation born after World War II. This cohort will soon reach middle age with its concomitant heart disease risks: the increase in age-specific prevalence will be accompanied by an increase in sheer size of the older age groups.

To make a quantitative analysis of the dynamics of the heart disease epidemic, we developed a simulation model for the most important manifestations of ischemic heart disease and congestive heart failure. Available information from the medical literature was combined with data obtained from mortality and morbid-

Luc Bonneux, Jan J. Barendregt, and Paul J. van der Maas are with the Department of Public Health, Karin Meeter is with the Department of Cardiology, Thoraxcenter, and Gouke J. Bonsel is with the Institute of Medical Technology Assessment, all at Erasmus University, Rotterdam, the Netherlands.

Requests for reprints should be sent to Luc Bonneux, MD, MSc, Department of Public Health, Erasmus University Rotterdam, P B 1738, 3000 DR Rotterdam, the Netherlands.

This paper was accepted June 11, 1993.

ity registers. We explored the future of heart disease morbidity in the Netherlands in different scenarios.

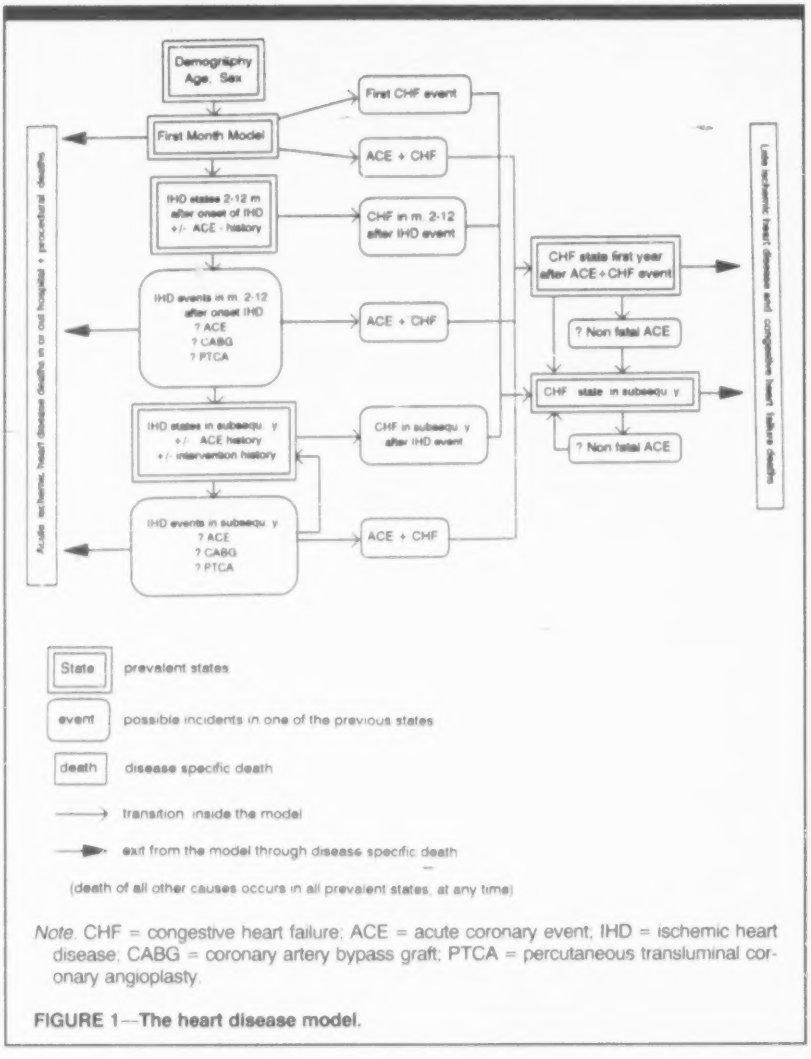
Methods

Model Structure

The deterministic state transition model used divides the population under study (here the entire Dutch population) into homogeneous subpopulations called states, defined by age, sex, and disease history. Each year, the population "ages" by 1 year and the states are updated. Transitions from one state to another are governed by transition probabilities, extracted from registries and literature. Transition probabilities are independent from the preceding states and depend only on the current state (Markov chain assumption).[12] The structure of the model resembles the Coronary Heart Disease Policy Model of Weinstein et al., but we have excluded cardiac arrest as a prevalent state and included congestive heart failure.[13]

The morbidity model consists of three main sectors: the first-month model, the ischemic heart disease model, and the congestive heart failure model. Persons developing heart disease, the rate of which is dependent on age and sex, move into the first-month model, which characterizes the initial event and its immediate sequelae during the first 30 days following the onset of disease. Persons in whom congestive heart failure forms the first heart disease event move directly into the appropriate model. Survivors of the first 30 days enter the ischemic heart disease model or the congestive heart failure model that describes the remaining 11 months of the first year and all subsequent years. Persons with ischemic heart disease can move into the congestive heart failure model, but not vice versa. Death from heart disease is modeled through the specific transition probabilities of the model. In all prevalent states, persons are subjected to an additional risk of death from all other causes, corrected for heart disease, that is based on national life tables.[14]

The first-month model covers the month immediately after the initial cardiac disease event (Figure 1). This event may be angina pectoris, an acute coronary event (an acute myocardial infarction or cardiac arrest), or congestive heart failure. Persons suffering a first acute coronary event may die before reaching a health care facility (deaths before admission).


Demography
Age, Sex
First Month Model
First CHF event
ACE + CHF
IHD states 2-12 m after onset of IHD +/- ACE - history
CHF in m. 2-12 after IHD event
IHD events in m. 2-12 after onset IHD ? ACE ? CABG ? PTCA
ACE + CHF
IHD states in subsequ. y +/- ACE history +/- intervention history
CHF in subsequ. y after IHD event
IHD events in subsequ. y ? ACE ? CABG ? PTCA
ACE + CHF
CHF state first year after ACE+CHF event
? Non fatal ACE
CHF state in subsequ. y
? Non fatal ACE
Acute ischemic heart disease deaths in or out hospital + precardial deaths
Late ischemic heart disease and congestive heart failure deaths
State
prevalent states
event
possible incidents in one of the previous states
death
disease specific death
transition inside the model
exit from the model through disease specific death
(death of all other causes occurs in all prevalent states, at any time)

Note. CHF = congestive heart failure; ACE = acute coronary event; IHD = ischemic heart disease; CABG = coronary artery bypass graft; PTCA = percutaneous transluminal coronary angioplasty.

FIGURE 1—The heart disease model.

Once taken into care, they are subjected to a further death risk (deaths in hospital). If the person survives the coronary event, the pump function of the heart may be permanently impaired. Patients with heart failure after an acute myocardial infarction enter the congestive heart failure model; the others move into the ischemic heart disease model. Persons with a first diagnosis of uncomplicated angina pectoris or congestive heart failure proceed to the ischemic heart disease and congestive heart failure models, respectively.

On entry, the ischemic heart disease model distinguishes between angina pectoris and acute coronary event states. Conditional on these states, patients face possible new events in the 2 to 12 months after the first incident. These events range from acute coronary events to coronary artery bypass grafts to percutaneous transluminal coronary angioplasties to all possible combinations of these. Each combination of events in a state has a probability of sudden death from ischemic heart disease dependent on the current event, the current state (and therefore the ischemic heart disease history), and the patient's age and sex. The patients also risk heart failure as a sequel of an acute coronary event ("ACE + CHF" in the figure) or as a consequence of chronic ischemic damage; these patients move into the congestive heart failure model. If none of the above occurs, patients will remain in the ischemic heart disease model.

After this first year, patients remaining in the ischemic heart disease model are reclassified into four new states, conditional on previous disease history. These states are described by the presence or absence of an acute coronary event or an intervention. These patients face the same events as in the previous year. At the end of each subsequent model year, persons

TABLE 1—Transition Probabilities for the Age Group 60 through 64 Years

	Men, %	Women, %
First incidence in the total population		
ACE	0.80	0.26
AP	0.47	0.30
CHF	0.20	0.09
Yearly probability of an ACE in the diseased population		
In the population with a history of ACE		
Before no intervention in y 1	5.3	5.3
Before PTCA in y 1	7.1	7.1
Before CABG in y 1	10.7	10.7
After CHF after ACE in y 1	20.5	19.4
After no intervention in y > 1	2.8	2.8
After any intervention in y > 1	4.3	4.3
After CHF in y > 1	7.3	6.2
In the population with a history of uncomplicated AP		
Before no intervention in y 1	1.5	0.6
Before PTCA in y 1	3.7	3.7
Before CABG in y 1	5.5	5.5
After no intervention in y > 1	1.5	0.7
After any intervention in y > 1	2.2	2.2
Probability of CHF in the diseased population		
After ACE in mo 1	8.5	11.1
After ACE in mo > 1	2.5	2.5
After AP	1.0	0.5
Intervention rates		
CABG		
In y 1 after ACE	11.3	5.9
In y > 1 after ACE	2.3	1.2
In y 1 after AP	15.9	8.8
In y > 1 after AP	3.2	0.6
PTCA		
In y 1 after ACE	2.6	1.9
In y > 1 after ACE	0.5	0.4
In y 1 after AP	3.9	2.9
In y > 1 after AP	0.8	0.8
Mortality rates		
Heart disease		
After ACE in mo 1		
First ACE, before admission	17.4	18.0
First ACE, inside hospital	8.4	11.0
Recurrent ACE, before admission	24.0	24.8
Recurrent ACE, inside hospital	12.7	16.5
Recurrent ACE after CHF, before admission	47.1	44.0
Recurrent ACE after CHF, inside hospital	18.9	22.0
Operative mortality after first CABG	1.4	3.0
Operative mortality after recurrent CABG	3.4	3.4
Procedural mortality after PTCA	0.7	0.7
After CHF after ACE in mo 2–11	19.9	17.9
After CHF in y > 1	13.0	11.0
All other causes of death in all states	1.14	0.64

Note. ACE = acute coronary event (including acute myocardial infarction, unstable angina pectoris, and cardiac arrest); AP = uncomplicated angina pectoris; CHF = congestive heart failure; CABG = coronary artery bypass grafting; PTCA = percutaneous transluminal coronary angiography; in y > 1 = in the years subsequent to the first year after an event.

"age" 1 year and are reassigned to states that reflect their updated history.

Persons with heart failure after an acute myocardial infarction face a high risk of dying during the first year.[15] If they survive the first year, they move to the congestive heart failure state describing the subsequent years, where they will stay until death; this state is called "chronic" to distinguish it from the first critical year following such a complicated infarction. In these "chronic" years, the model assumes that mortality rates are lower. All other patients developing heart failure immediately enter the chronic congestive heart failure state.

Although nonfatal acute coronary events are possible in the congestive heart failure state, such events no longer effect any change in the assigned state. Patients in the congestive heart failure state face an annual probability of death due to cardiovascular causes that may be classified as acute myocardial infarction (ICD 410–411), chronic ischemic heart disease (ICD 412–414), congestive heart failure (ICD 428–429), or stroke (ICD 430–438).[16] We have incorporated stroke as a specific cause of death because the Framingham Heart Study showed that, because of common risk factors (hypertension and diabetes), congestive heart failure is associated with increased rates of stroke and death from stroke.[17]

Incidence and Recurrence Rates

Reliable and recent incidence data on heart disease are not available in the Netherlands. Incidence figures have been estimated by an iterative nonlinear procedure from observed numbers of events (acute myocardial infarctions, ischemic heart disease deaths, interventions) and externally estimated transition probabilities.[18] In this section we give a brief account of how they were derived. The transition probability matrix for a typical age group (60 through 64 years) is furnished in Table 1. The complete transition matrices and a more detailed account of references are available from the authors on request.

Total numbers of admissions for acute myocardial infarction (ICD 410), total numbers of hospital deaths, and total numbers of interventions by age and sex are available from the nationwide hospital register.[19–22] Patients with symptomatic infarctions who are not admitted to hospital (6%; in the Netherlands, this group is limited to the oldest age groups) are added.[23] A constant ratio for the probability of dying before hospital admission versus dying inside the hospital from an acute coronary event is assumed.[20,21,24,25]

Reported probabilities of recurrent acute coronary events vary widely, from 2% to 13% and more per year, depending on age, case mix, and selection criteria.[13,26–30] We have assumed that ischemic heart disease patients without congestive heart failure have a first annual probability of a recurrent acute coronary event of 0.062. This probability is consistent with the estimate of Weinstein et al. of 0.078, in which patients with congestive heart failure are included.[13,31] Recurrence rates are strongly age dependent. We assume the age distribution to be identical to the known age distribution of a first acute coronary event in the reference population. First acute coronary events after a history

of angina pectoris were modeled on the basis of data derived from the Framingham study, which showed that male patients with angina pectoris, who have suffered no previous acute myocardial infarction have a relative risk of 0.54 compared with male patients with a previous acute myocardial infarction.[32] Female patients with angina pectoris have a relative risk of 0.45 compared with male patients.[32]

Patients eligible for coronary artery bypass graft surgery run a risk for recurrent events that is twice as high as that for patients who are not eligible.[31,33–35] Coronary artery bypass graft surgery reduces hazard rates by 20%.[36,37] Lacking data on the effectiveness of percutaneous transluminal coronary angioplasty, we assumed the relative risk for recurrent acute coronary events to be 1.5, with a rate reduction of 10%.

Intervention Rates

Intervention rates are known to vary widely from country to country and even from hospital to hospital.[38,39] We have used age- and sex-specific nationwide output data and assumed constant proportional relations between states to estimate the numbers of interventions in each state. The risk of an intervention after uncomplicated angina pectoris vs the risk of an intervention after an acute coronary event is 1.4: 35% of patients with a history of an acute coronary event are potentially eligible for an intervention, compared with 50% in uncomplicated angina.[40,41] The risk of undergoing an intervention in the first year after an ischemic heart disease event vs the risk in subsequent years was taken to be 5.[42] The annual probabilities of having a subsequent intervention after a first intervention are 1.0% (coronary artery bypass graft) and 0.5% (percutaneous transluminal coronary angioplasty).[43] For every 10 successful percutaneous transluminal coronary angioplasties, another 4 are considered to be unsuccessful because of restenosis.[44] If both a coronary artery bypass graft and percutaneous transluminal coronary angioplasty occur in the same year in the same patient, the percutaneous transluminal coronary angioplasty is ignored.

Heart Failure

Heart failure due to ischemic heart disease is caused by acute events or by chronic ischemic damage. Eighteen percent of ischemic heart disease patients survive an acute myocardial infarction with irreversible heart failure as a sequel.[33,34] The distribution by age, sex, and previous ischemic heart disease history is assumed to be the same as the distribution of in-hospital mortality by age, sex, and previous ischemic heart disease history after an acute myocardial infarction. The annual probabilities of developing congestive heart failure, given a history of ischemic heart disease, were estimated from the Framingham Heart Study data.[17] We obtained the published annual event rates of 1.5% given an uncomplicated angina pectoris history and 3.0% given a history of an acute coronary event and adjusted to allow for the number of acute cases caused by an acute coronary event. The age distribution was calculated by applying relative risks of 3.0 (angina pectoris) and 5.0 (acute myocardial infarction) to the age-specific incidence data taken from the Framingham study.[17]

The incidence rates of congestive heart failure for persons without a history of ischemic heart disease are calculated by assuming that persons in whom congestive heart failure is the certified cause of death survived their first diagnosis of congestive heart failure for 5 (males) to 7 (females) years.[11,12,15,17]

Heart Disease Mortality

The total age- and sex-specific in-hospital mortality rates following acute coronary events and interventions are known in the Netherlands.[20,22] The relative risk of dying after a subsequent vs a first acute coronary event was taken to be 1.5.[13,28,45] Publications are available in which the figures on mortality due to a subsequent vs a first coronary artery bypass graft operation are given.[46] The relative risk of death due to percutaneous transluminal coronary angioplasty vs coronary artery bypass graft surgery is 0.5.[47,48] Procedural mortality due to a subsequent or a first percutaneous transluminal coronary angioplasty is taken to be the same.

The risk of death during the first year after acute myocardial infarction followed by congestive heart failure was assumed to be 25%; the distribution by age was constructed by exponential interpolation and extrapolation of data taken from Ahnve et al.[34] The annual mortality risk of heart failure during the subsequent years is assumed to be constant: 10% for females and 13% for males.[15,17,33,34] Of all deaths caused directly or indirectly by heart failure, 25% (female) and 37% (male) are attributable to recurring ischemic heart events and another 25% (males) and 35% (females) are attributable to stroke (and hence to noncardiac disease).[17] The figures for nonfatal acute coronary events in persons with heart failure are estimated by assuming a case fatality of 60%.

All persons in all states are subject to the risk of dying of other diseases, adjusted to allow for the causes of death under study and based on Dutch life tables.

Results

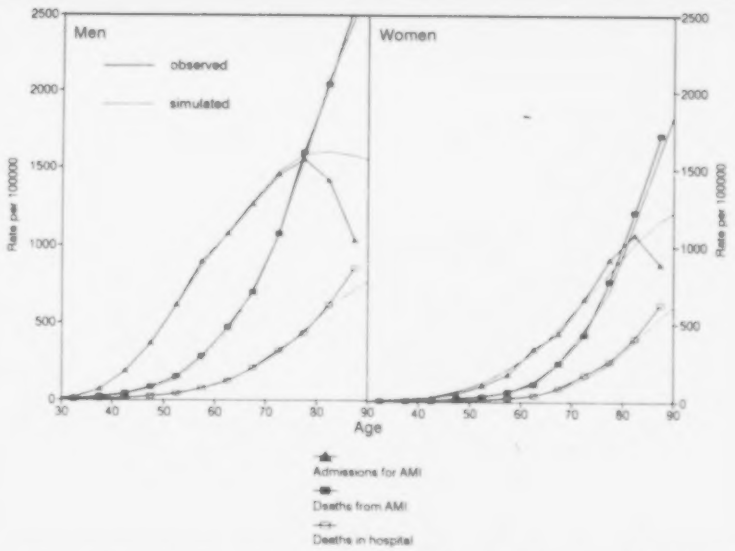
Calculated Estimates for 1985

Figure 2 shows the observed and simulated events of acute myocardial infarction: total admissions and deaths in and outside the hospital, by age and sex. The model closely reproduces the observed data. The gap between simulated and observed rates in the older age groups represents persons with recognized acute myocardial infarctions who are not admitted to a hospital. Calculated prevalences of heart disease by age and sex are shown in Figure 3. Heart disease is more concentrated in the oldest age groups in women, whereas the burden of morbidity in the male population is already high at middle age. The leveling off and decrease by age in the prevalence of clinical ischemic heart disease in men shown in the figure is caused by the steep rise in the incidence of congestive heart failure and sudden cardiac death, reducing clinical ischemic heart disease incidence. Prevalence of all heart disease reaches 25% in the age groups 70 through 74 years (men) and 80 through 84 years (women). Women suffer relatively more often than men from angina pectoris and congestive heart failure, but less often from severe ischemic heart disease. This pattern has been confirmed time and again in many surveys.[32,49,50]

Projections for 1985–2010

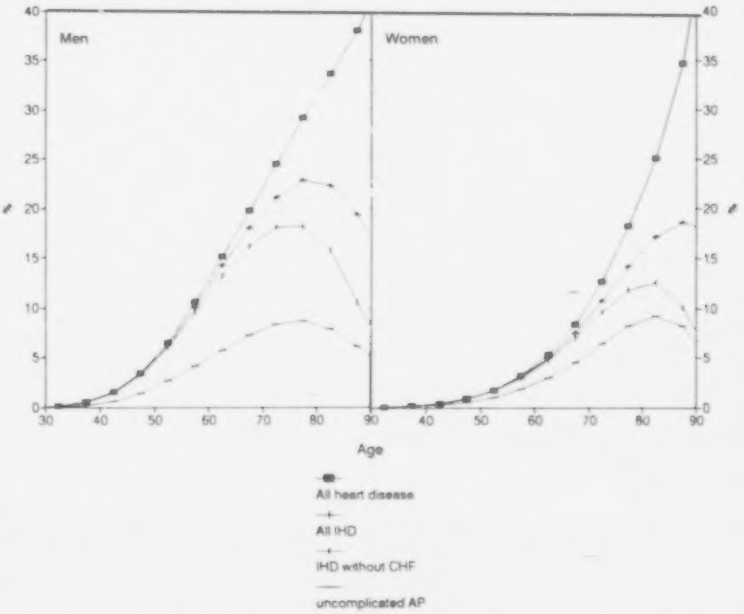
Disease-specific mortality is the direct result of incidence and survival. If two of these elements are known, the third can be calculated. Because the decline in ischemic heart disease mortality that has been observed since 1975 is indisputable, we postulated an ongoing decline of this mortality into the near future. We provide three plausible scenarios, based on varying assumptions of further decreasing incidence rates and improving prognoses (see Table 2).

The reference scenario, based on the Framingham Heart Study, supposes that one third of the mortality decline is explained by a decline in incidence (−1.0% per year) and that two thirds is explained



Note. "Admissions for AMI" are the rates of patients discharged with a primary diagnosis of acute myocardial infarction (ICD 410). "Deaths in hospital" are the rates of patients with the same diagnosis dying in hospital.[20] "Deaths from AMI" are the total rates of death from acute myocardial infarction (ICD 410).[21] The solid lines with symbols show observed numbers; the dotted lines represent the simulation by the heart disease model.

FIGURE 2—Observed and simulated events in 1985.



Note. "Uncomplicated AP" is defined as angina pectoris only, without a history of acute myocardial infarction. "IHD without CHF" describes all patients with a history of ischemic heart disease but without congestive heart failure. "All IHD" comprises all ischemic heart disease, including patients who developed congestive heart failure after ischemic heart disease. "All heart disease" summarizes all three conditions. The difference between all heart disease and all ischemic heart disease shows the prevalence of congestive heart failure.

FIGURE 3—Cumulative prevalences of heart disease, by age and sex, the Netherlands, 1985.

by an increase in survival.[3,4,8,9,26,51] In an extreme low-incidence-change scenario, we assume that the incidence decrease is even less: −0.5% per year. This is the magnitude of change predicted by a risk factor model, based on changes in risk factor prevalences in the Netherlands.[5] In an extreme high-incidence-change scenario, we assume a high decrease in incidence (−1.7% per year).

We have projected these trends over the period 1985 through 2010 to juxtapose the effects of demographic and epidemiologic change.

The results are summarized in Figures 4 through 6 and Table 3. Figure 4 shows how the decreasing incidence, together with the increase in survival, results in lower prevalence rates in the young and increasing prevalence rates in older age groups. In the youngest (30- to 39-year-old) age group, the decrease will be about 25%; in the oldest old (85 years and older), prevalence will increase by 20% for men and 10% for women. Superimposing the population projections on these figures results in a completely different picture, which is shown in the same figure. The movement of the postwar baby boom generation into middle age and the sharp increase in the number of the very elderly is clearly revealed. Together with the projected increase in prevalence rates, this demographic change will result in a steep climb in the number of heart disease patients in the oldest age groups. Hence, while epidemiological changes in cardiovascular disease will result in an increase of 10% to 20% in prevalence rates for the oldest old (85 years and older), the absolute number of patients will more than double.

Figures 5 and 6 show the projected changes in age-specific prevalences for different types of ischemic heart disease; Figure 5 shows the relative changes in rates and Figure 6 shows the absolute changes in numbers. The change in age-specific rates for all ischemic heart disease combined is the same as in Figure 4.

Figure 5 shows a relative increase of severe ischemic heart disease. Milder stages of the disease act as a risk factor for the more severe stages. If survival increases in all disease stages, effects will culminate in the most serious stages: as more patients stay alive with ischemic heart disease, more will have the opportunity to develop congestive heart failure. Figure 6 shows the changes in absolute terms: the relative decrease of prevalence at younger ages has hardly any impact because of the low number of patients with

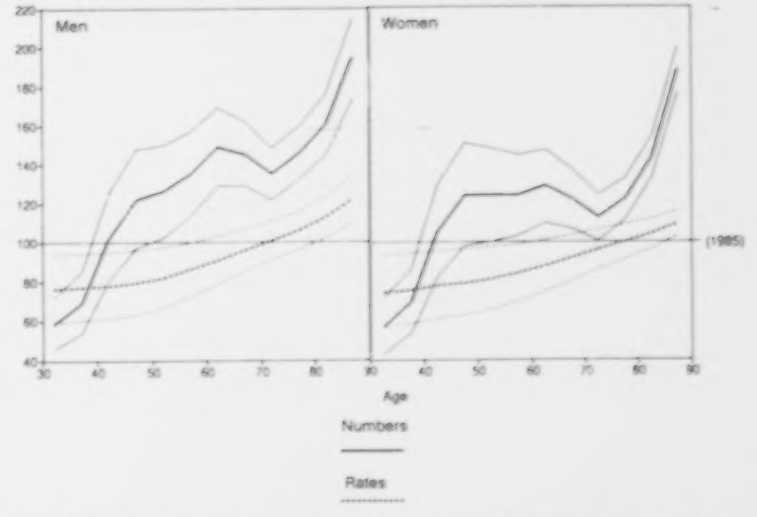
ischemic heart disease at early ages. However, the combination of aging and increasing survival of heart disease leads to a flood of middle-aged men with ischemic heart disease followed by a slower wave of elderly persons with heart failure.

Table 3 shows expected changes in age-standardized rates, using the European Standard Population as the standard, and in absolute numbers, combining the rates with the expected population dynamics in the Netherlands. The mortality decline (observed between 1975 and 1990 and projected over the next 20 years) is so strong that it will be evident even when expressed in absolute numbers. Although the assumed decline in incidence of first myocardial infarctions, which accounts for one third of the decline in mortality, is too small to counteract the aging of the Dutch population, the rate of all myocardial infarctions together will nevertheless decrease substantially. This decrease is the result of the combined decline in incidence and recurrence; the better survival, on the other hand, through which the number of patients at higher risk of recurrent infarctions is increased, is not strong enough to counteract this decline. Because more patients will stay alive and at risk for interventions and reinterventions and because of the aging of the population, the need for coronary artery bypass graft surgery and percutaneous transluminal coronary angioplasty will increase by 1% per year (range: 0.4%–1.6%). The age-standardized prevalence ratios for milder ischemic heart disease tend to decrease, but those for ischemic heart disease complicated by congestive heart failure increase strongly: in absolute numbers, a growth of more than 70% is to be expected.

TABLE 2—Annual Percentage Changes Postulated in Three Scenarios for the Period 1985 through 2010

	Men	Women
	Reference (Extremes)	Reference (Extremes)
Incidence	−1.1 (−1.7, −0.6)	−0.9 (−1.4, −0.5)
Mortality before admission	−1.2 (−0.9, −1.4)	−0.7 (−0.5, −0.8)
Mortality within first month after admission	−2.1 (−1.6, −2.5)	−1.3 (−1.0, −1.6)
Late mortality	−1.8 (−1.4, −2.1)	−1.6 (−1.4, −2.1)
Recurrence	−3.1 (−2.3, −4.7)	−1.9 (−1.4, −2.9)
Ischemic heart disease mortality	−2.4	−2.0

Note. It is assumed that the mortality decline observed between 1975 and 1990 remains equal in all scenarios.



Note. Dashed lines show the expected index of change in prevalence rates between 1985 and 2010 (prevalence in 2010 divided by prevalence in 1985 multiplied by 100). Solid lines indicate the change in absolute numbers, showing the force of demography. The heavier lines indicate the reference scenario; the thinner lines indicate the extreme scenarios.

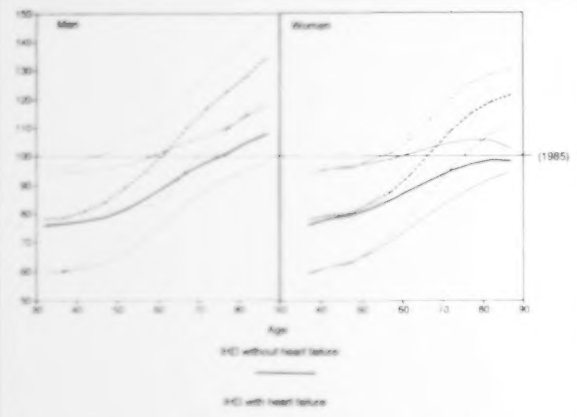
FIGURE 4—Expected changes in absolute and relative prevalence of patients with ischemic heart disease, the Netherlands, 2010 compared with 1985.

Discussion

Simulation modeling provides quantitative insight into epidemiological and public health dynamics and potential future developments. The information needed for these models can never be supplied by a single data source. Combining data from different sources is therefore appropriate, and the use of state-transition models provides a simple and comprehensible means of doing so. The methods used are relatively easy to understand. The parameters of the model are meaningful to physicians and the results are interpretable by health care providers, who are the principal users of these results.

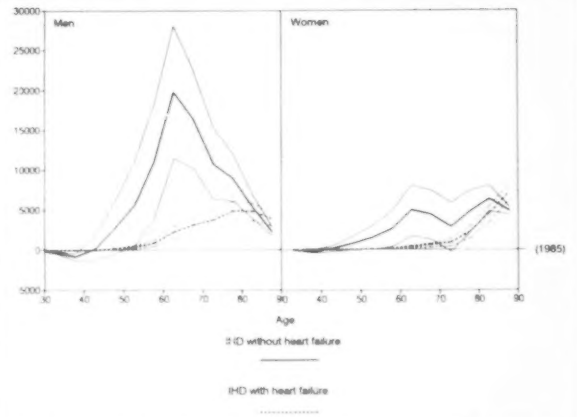
Nevertheless, to interpret results correctly, the limitations of this kind of modeling should be understood. A number of simplifying assumptions are always necessary to keep the model, and more particularly the number of parameters to be estimated, manageable. We included certain disease manifestations and excluded others. Unlike Weinstein et al., we excluded cardiac arrest as a prevalent state but included congestive heart failure. Cardiac arrest is highly lethal, and no major error is introduced by mixing the history of those surviving an acute myocardial infarction with that of a few survivors of cardiac arrest. On the other hand, the inclusion of congestive heart failure takes into account an important cause of chronic and severe morbidity. Modeling the dynamics of and the relations between these different types and levels of morbidity should result in a more realistic estimate of future health care needs.

We used registered admissions for acute myocardial infarction (ICD 410), deaths from ischemic heart disease (ICD 410–414) and congestive heart failure (ICD 428–429), and transition probabilities extracted from multiple sources to estimate incidences. This complicates the assessment of validity, which is the most important problem of this or any other model with such complexity. The model output reflects state-of-the-art clinical evidence, closely simulating present hospitalization, intervention, and mortality rates (see Fig-



Note. Solid lines indicate the relative changes in number of ischemic heart disease patients without heart failure; dashed lines indicate those with heart failure (age-specific prevalences in 2010 divided by the same prevalences in 1985 multiplied by 100). The heavier lines indicate the reference scenario; the thinner lines indicate the extreme scenarios.

FIGURE 5—Expected relative changes in prevalence rates of ischemic heart disease, the Netherlands, 2010 compared with 1985.



Note. Solid lines indicate the absolute changes in number of ischemic heart disease patients without heart failure; dashed lines indicate those with heart failure (age-specific numbers of patients in 2010 minus the same numbers in 1985). The heavier lines indicate the reference scenario; the thinner lines indicate the extreme scenarios.

FIGURE 6—Expected absolute changes in numbers of patients with ischemic heart disease, the Netherlands, 2010 compared with 1985.

TABLE 3—Age-Standardized Rates[a] and Absolute Numbers of Admissions, Interventions, Patients, and Deaths due to Ischemic Heart Disease (IHD) and Congestive Heart Failure (CHF) in the Netherlands, in 1985 and the Projected Index of Change (1985 = 100) in 2010

	1985		Index of Change in 2010 (1985 = 100)	
	Rate	Number	Rate	Number
Men				
ACE incidence	267	16 438	76 (65, 87)	114 (97, 130)
First admissions	214	13 150	80 (67, 93)	119 (100, 139)
Admissions for ACE	352	21 687	67 (60, 73)	100 (90, 108)
CABG	96	5 797	84 (72, 90)	128 (110, 147)
History of AP	1726	106 968	91 (79, 104)	136 (119, 156)
History of ACE	2087	129 149	90 (79, 102)	136 (118, 153)
CHF with history of IHD	575	35 761	118 (104, 131)	170 (149, 189)
CHF without history of IHD	697	41 323	94 (88, 126)	144 (133, 163)
Deaths coded AMI	194	12 045	54[b]	80[b]
Deaths coded chronic IHD or CHF	106	6 422	65[b]	96[b]
Women				
ACE incidence	109	9 507	80 (71, 89)	115 (105, 126)
First admissions	81	6 883	83 (73, 94)	118 (105, 131)
Admissions for ACE	122	10 529	73 (66, 78)	105 (96, 111)
CABG	18	1 302	84 (72, 97)	115 (98, 132)
History of AP	1088	94 635	89 (78, 99)	123 (110, 136)
History of ACE	547	45 195	94 (82, 106)	130 (114, 146)
CHF with history of IHD	294	29 251	115 (103, 125)	169 (153, 183)
CHF without history of IHD	509	50 729	84 (77, 104)	130 (120, 146)
Deaths coded AMI	82	7 889	63[b]	96[b]
Deaths coded chronic IHD or CHF	60	6 036	59[b]	90[b]

Note. The expected rates and numbers can be calculated by multiplying the figures for 1985 by the index of change in 2010 divided by 100. The upper and lower bounds (in parentheses) represent the results of scenarios assuming extremely high or low incidence changes but leading to the same mortality changes as a result of extremely low or high survival changes (see Table 2). ACE = acute coronary event; CABG = coronary artery bypass graft; AP = angina pectoris; AMI = acute myocardial infarction.

[a]Per 100 000; the European Standard Population is used as the standard.

[b]In the calculation of upper and lower bounds, the mortality decline is assumed to remain equal.

ure 2). Its estimated prevalences of clinical disease coincide closely with empirical observations.[49,50,52]

Our forecasts are based on mortality projections. For health care planning purposes, projections of incidence based on past risk factor prevalences are less reliable. First, prevalences of all known risk factors for heart disease are not available; prevalence numbers that are available are often highly inconsistent between studies. For example, estimates of the prevalence of hypertension in Dutchmen aged 50 through 59 years vary from 14.4% to 67%.[5,52] Second, not all risk factors are known or observable; projections based only on known risk factors are bound to underestimate future trends. The assumption that mortality will continue to decline is uncertain, but not implausible. It supposes a progressive change in risk factor distributions, a more widespread use of effective health care interventions in the future, and ongoing technological advances.[3,4,8,9] The level of ischemic heart disease mortality expected in 2010 is comparable to that of 1950 (making allowances for changes in codification). Given the absence of effective therapies in 1950, this is not an extreme result. The reference scenario supposes a decrease in incidence, explaining one third of the mortality decline, based on the Framingham study.[4] Other assumptions about incidence decreases and survival increases within extreme ranges are shown in the alternative

scenarios.[3,4,8] These assumptions influence the absolute magnitude of the expected changes, but not the direction.

Declining incidence rates mean that the onset of disease is postponed but not halted. The delay in occurrence causes a reduction of morbidity in the young, which, however, is canceled out with age. Mortality is deferred to an even greater extent, owing to a combination of the decrease in incidence and improved prognosis. This mortality decrease will be traded off for a steep increase of more severe morbidity in the old.

Changes in demography will prevent a decrease of new cases of heart disease.[13] However, while age-specific incidence rates show a decline, acute myocardial infarction rates will drop more steeply, owing to a reduction in both incidence and recurrence rates; the increasing prevalences are insufficient to cancel these reductions. The reference scenario predicts a decline in acute myocardial infarction rates that will be sufficiently strong to offset population aging; this development has been observed in the Netherlands since 1985.[19–21] But although disease incidence rates are declining, intervention rates will remain unchanged, even assuming that intervention rates per group of patients in a certain disease state remain constant. The reason is that more patients will survive and, on average, survival is longer, resulting in more people in disease states who are eligible for invasive coronary revascularization procedures.

The overall picture emerging from the heart disease scenarios is one of transition from an acute to a more chronic disease: event rates are declining steeply, but the postponed mortality is traded off for increased morbidity and severity of disease among the elderly. This pattern can be observed in hospital admission rates in the Netherlands, in data from the National Health Interview Survey in the United States, and in data from the National Health Survey of Japan.[19,53] All of these data show increasing prevalences of ischemic heart disease. As a consequence, overall health care needs will continue to rise, regardless of declining mortality and incidence.

These calculations, based on empirical observations, shed a new light on Fries' compression of morbidity theory.[54] The compression of morbidity theory hinges on the proposition that the incidence of disease will be postponed more rapidly than mortality. For heart disease, the evidence suggests that the decline of lethality expands the number of years spent with heart disease morbidity, regardless of the incidence postponement. Healthy life expectancy is extended, as expected by Fries and assumed by the scenarios, but contrary to his prediction, total life expectancy is extended even further. Among the middle-aged and elderly, acute heart disease lethality in the Netherlands is high, increasing sharply with age. Obviously, here would appear to be a serious potential for reduction. In 1985, for example, the model calculated that 41% of the 70-year-old men died within 4 weeks of their first acute coronary event.[20–24] In our scenarios 32% (range: 30%–34%) are predicted to die within the same period in the year 2010, which in our view remains a conservative estimate. Because heart disease is concentrated in the older age groups, even small differences in the rates of change in incidence and mortality can have major consequences, particularly now that pushing back mortality has been shown to be counterbalanced by an increase in the more severe stages of heart disease. With short-term case-fatality subsequent to an acute event as high as it is, lowering these rates will remain an important goal of health care intervention. Should this aim be successful, the resulting increase in the number of survivors will boost heart disease morbidity and lengthen life expectancy with disease. Only if health care interventions aimed at reducing mortality subsequent to the event should prove a waste of time and money would no expansion of morbidity be expected. □

Acknowledgments

This study was funded by the Ministry of Welfare, Health and Cultural Affairs, The Netherlands.

We wish to thank B. Michel (from the Department of Public Health), F. van den Burg (medical student), L. Niessen, K. Gribling-Laird, and the anonymous referees for their comments, help, and editing.

Members of the Technology Assessment Methods project team are J. J. Barendregt, E. van Beeck, R. Boer, L. Bonneux, G. J. Bonsel, L. Gunning-Schepers, J. D. F. Habbema, B. A. van Hout, B. M. van Ineveld, M. A. Koopmanschap, C. W. N. Looman, J. Lubbe, P. J. van der Maas, J. P. Mackenbach, G. J. Oortmarssen, L. van Roijen, and F. H. H. Rutten.

References

1. Koopmanschap MA, van Roijen L, Bonneux L. *Costs of Illness in the Netherlands* [in Dutch]. Rotterdam, the Netherlands: Department of Public Health and Social Medicine, Institute of Medical Technology Assessment; 1991. Report MGZ 91.03.
2. Uemura K, Pisa Z. Trends in cardiovascular disease mortality in industrialised countries. *World Health Stat Q.* 1988;41:155–168.
3. Sytkowsky PA, Kannel WB, D'Agostinho RB. Changes in risk factors and the decline in mortality from cardiovascular disease. *N Engl J Med.* 1990;322:1635–1641.
4. Sigfusson N, Sigvaldason H, Steingrimsdottir L, et al. Decline in ischemic heart disease in Iceland and change in risk factor levels. *BMJ.* 1991;79:366–371.
5. Gunning-Schepers L. The health benefits of prevention: a simulation approach. *Health Policy.* 1989;12/1–2 (special issue): 93–129.
6. Graboys TB, Biegelsen B, Lampert S, Blatt CM, Lown B. Results of a second-opinion trial among patients recommended for coronary angiography. *JAMA.* 1992; 268:2537–2540.
7. Mackenbach JP, Looman CWM, Kunst AE, Habbema JDF, van der Maas PJ. Post-1950 mortality trends and medical care: gains in life expectancy due to declines in mortality from conditions amenable to medical intervention in the Netherlands. *Soc Sci Med.* 1988;27:889–894.
8. Goldman L, Cook EF. The decline in ischemic heart disease mortality rates. *Ann Intern Med.* 1984;101:825–836.
9. Moss AJ, Benhorin J. Prognosis and management after a first myocardial infarction. *N Engl J Med.* 1990;322:743–753.
10. SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med.* 1991;325: 293–302.
11. Cohn JN, Johnson G, Ziesche S, et al. A comparison of enalapril with hydralazine-isosorbide dinitrate in the treatment of chronic congestive heart failure. *N Engl J Med.* 1991;325:303–310.
12. Manton KG, Stallard E. *Chronic Disease Modelling.* London, England: Charles Griffins & Co; 1988:207–210.
13. Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Forecasting coronary heart disease incidence, mortality and cost: the Coronary Heart Disease Policy Model. *Am J Public Health.* 1987;77:1417–1426.
14. *Monthly Population Statistics.* Voorburg, the Netherlands: Central Bureau of Statistics; 1986(3):20–28.
15. Smith WM. Epidemiology of congestive heart failure. *Am J Cardiol.* 1985;55:3A–8A.
16. *International Classification of Diseases, Injuries and Causes of Death.* 9th revision. Geneva, Switzerland: World Health Organization; 1977.
17. Kannel WB. Epidemiological aspects of heart failure. *Cardiol Clin.* 1989;7:1–9.
18. Press W, Flannery BP, Teukolsky SA, Vetterling T. *The Art of Scientific Computing.* Cambridge, England: Cambridge University Press; 1990.
19. Hoogendoorn DE. Atherosclerotic diseases of the heart in the hospital and in the vital statistics [in Dutch]. *Ned Tijdschr Geneeskd.* 1985;129:1827–1833.
20. Hoogendoorn D. Some notes on the current situation regarding the epidemic of acute myocardial infarction [in Dutch]. *Ned Tijdschr Geneeskd.* 1990;134:592–595.

21. *Files from the National Medical Registration concerning Admissions and Interventions in 1988.* Utrecht, the Netherlands: LMR (Information Center for Health Care), 1990.
22. Hoogendoorn D. Frequencies and operative mortality of interventions on the heart and the large intrathoracic vessels, I: direct revascularisation of the coronary arteries [in Dutch]. *Ned Tijdschr Geneeskd.* 1989; 133:2445–2448.
23. Fracheboud J. *Coronary Care or Home Care? A Descriptive Study of Home Care for Patients with an Acute Myocardial Infarction* [in Dutch]. Utrecht, the Netherlands: Nederlands Instituut Voor onderzoek van de Eerste Lijnsgezondheidszorg (NIVEL); 1987.
24. Lubsen J, van der Does E, Pool J. Incidence of acute myocardial infarction and sudden death. *Hart Bull.* 1976;7:107–113.
25. Goldman L, Cook F, Hashimoto B, Stone P, Muller J, Loscalzo A. Evidence that hospital care for acute myocardial infarction has not contributed to the decline in coronary mortality between 1973–74 and 1978–79. *Circulation.* 1982;65:936–942.
26. Ericsson CG, Erhardt L, Rehnqvist N. Two-year survival after myocardial infarction. *J Intern Med.* 1990;227:195–199.
27. Marcus FI, Friday K, McCans J, et al. Age related prognosis after acute myocardial infarction (the Multicenter Diltiazem Postinfarction Trial Research Group). *Am J Cardiol.* 1990;65:559–566.
28. Benhorin J, Moss AJ, Oakes D, and the Multicenter Diltiazem Postinfarction Trial Research Group. Prognostic significance of nonfatal myocardial reinfarction. *J Am Coll Cardiol.* 1990;15:253–258.
29. Multicenter Diltiazem Postinfarction Trial Research Group. The effect of diltiazem on mortality and reinfarction after myocardial infarction. *N Engl J Med.* 1988;319:385–392.
30. Norwegian Multicenter Study Group. Timolol induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. *N Engl J Med.* 1981; 304:801–807.
31. DeBusk RF, Kraemer HC, Nash E, Berger WE III, Lew H. Stepwise risk stratification soon after acute myocardial infarction. *Am J Cardiol.* 1983;52:1161–1166.
32. Kannel WB, Feinleib M. Natural history of angina pectoris in the Framingham study. *Am J Cardiol.* 1972;29:154–163.
33. DeBusk RF, Blomqvist G, Kouchoukos NT, et al. Identification and treatment of low-risk patients after acute myocardial infarction and coronary artery bypass graft surgery. *N Engl J Med.* 1986;314:161–166.
34. Ahnve S, Gilpin E, Dittrich H, et al. First myocardial infarction: age and ejection fraction identify a low-risk group. *Am Heart J.* 1988;116:925–932.
35. Sanz G, Castañer A, Betriu A, et al. Determinants of prognosis in survivors of myocardial infarction. *N Engl J Med.* 1982; 306:1065–1070.
36. European Coronary Study Group. Long term results of a prospective randomised study of coronary artery bypass surgery in stable angina pectoris. *Lancet.* 1982;ii: 1173–1180.
37. Coronary Artery Surgery Study principal investigators and associates. Coronary Artery Surgery Study: a randomized trial of coronary artery bypass surgery. *Circulation.* 1983;68:939–950.
38. Brook RH, Kosecoff JB, Park RE, Chassin MR, Winslow CM, Hampton JR. Diagnosis and treatment of coronary artery disease: comparisons of doctors' attitudes in the USA and the UK. *Lancet.* 1988;i:750–753.
39. Winslow CM, Kosecoff JB, Chassin M, Kanouse DE, Brook RH. The appropriateness of performing coronary artery bypass surgery. *JAMA.* 1988;260:505–509.
40. Weintraub WS, Jones EL, King SB, et al. Changing use of coronary angioplasty and coronary bypass surgery in the treatment of chronic coronary artery disease. *Am J Cardiol.* 1990;65:183–188.
41. Kaul S, Lilly DR, Gascho JA, et al. Prognostic utility of the exercise thallium-201 test in ambulatory patients with chest pain: comparison with cardiac catheterization. *Circulation.* 1988;77:745–758.
42. Parisi AF, Peduzzi P, Detre K, et al. Characteristics and outcome of medical nonadherers in the Veterans Administration cooperative study of coronary artery surgery. *Am J Cardiol.* 1984;53:23–28.
43. Veldkamp RF, Baartman GJ, van Domburg R, Thijssen JGP, Bos E, Meeter K. Survival 11 years after an aortocoronary bypass operation [in Dutch]. *Ned Tijdschr Geneeskd.* 1991;135:1229–1233.
44. Glazier JJ, Varicicchione TR, Ryan TJ, et al. Outcome in patients with recurrent stenosis after percutaneous transluminal balloon angioplasty. *Br Heart J.* 1989;61:485–488.
45. Fioretti P, Tijssen JGP, Azar AJ, et al. Prognostic value of predischarge 12 lead ECG after myocardial infarction compared with other routine variables. *Br Heart J.* 1987;57:306–312.
46. Laird-Meeter K, van Domburg R, van den Brand MJBM, Lubsen J, Bos E, Hugenholtz PG. Repeat interventions after initial aorto-coronary bypass surgery and its results [in Dutch]. *Ned Tijdschr Geneeskd.* 1988;132:2316–2320.
47. Detre K, Holubkow R, Kelsey S, et al. One year follow-up results of the 1985–1986 National Heart, Lung, and Blood Institute's PTCA registry. *Circulation.* 1989;80:421–428.
48. Epstein SE, Palmeri ST, Patterson RE. Evaluation of patients after acute myocardial infarction. *N Engl J Med.* 1982;307: 1487–1492.
49. Anonymous. Epidemiology of chest pain and angina pectoris, with special references to treatment needs. *Acta Med Scand Suppl.* 1984;682:1–120.
50. Reunanen A, Aroma A, Pyörälä K, Punsar S, Maatela J, Knekt P. The social insurance institution's coronary heart disease study. *Acta Med Scand Suppl.* 1983;673:1–120.
51. de Cock CC, Visser FC, van Eenige MJ. Prognosis after an infarction: short- and long-term follow-up data [in Dutch]. *Ned Tijdsch Cardiol.* 1991;5:96–104.
52. *Morbidity Figures from General Practice.* Nijmegen, the Netherlands: Nijmeegs Universitair Huisartsen Instituut; 1985.
53. Uemura K. International trends in cardiovascular diseases in the elderly. *Eur Heart J.* 1988;9(suppl D):1–8.
54. Fries JF. Aging, natural death and the compression of morbidity. *N Engl J Med.* 1980; 303:130–135.

Self-Reported Physical Activity in a Rural County: A New York County Health Census

Charles B. Eaton, MD, MS, Anne N. Nafziger, MD, MHS, David S. Strogatz, PhD, and Thomas A. Pearson, MD, PhD, MPH

ABSTRACT

Objectives. Few studies have described physical activity in rural populations. This study describes the frequency, types, and correlates of physical activity in 29 304 free-living adults in a rural county in New York State.

Methods. Self-reported responses about regular physical activity (maintained long enough to work up a sweat) were analyzed from a private household census of Otsego County with an 86.6% response rate.

Results. This survey categorized 46.2% of county residents as sedentary. Walking, the most frequent choice of activity (62% of the women, 36% of the men), increased in frequency with age of respondents whereas cycling, jogging, aerobics, team sports, and swimming (listed in rank order of frequency) generally tended to decrease in frequency with age. Farmers demonstrated an increased amount of "sweat activity" compared with persons in most other occupations.

Conclusions. This descriptive study of physical activity in a rural county shows that sedentary lifestyle is of high prevalence. The high frequency of walking and the gender differences in both the levels and choice of activity suggest that further research and public policy recommendations focus on these issues. (*Am J Public Health.* 1994;84:29–32)

Introduction

For public health programs to be targeted appropriately, information on the types and frequency of physical activity in various groups, such as rural populations, needs to be obtained. Most previous community studies have been performed in urban or suburban settings, not in rural ones.[1–4] Larger national studies have been conducted,[5–11] but only two[5,6] have been analyzed comparing urban vs rural location. Data on risk factors such as physical activity may help explain the usual advantage of rural areas in urban-rural comparisons of coronary heart disease mortality and trends in coronary heart disease mortality over time.[12–15] This paper describes the frequency, types, and correlates of physical activity in a rural population in New York State.

Methods

Door-to-door collection and assistance with completing a private census form led to a 86.6% level of participation rate representing 17 144 out of 19 800 households. The section of the census form giving data about physical activity, self-reported height and weight, and standard sociodemographic information was completed by 33 428 individuals aged 17 years or older. To assess levels of physical activity, respondents were asked, "At least once a week, do you engage in any regular activity like brisk walking, jogging, bicycling, etc. long enough to work up a sweat? (No, Yes) If yes, how many times a week? _____ Activity _____."

This question, which was derived from Paffenbarger et al.'s[16] original physical activity questionnaire, has been validated by Siconolfi et al.,[17] Kohl et al.,[18] and Washburn et al.[19] and shows correlations with measures of cardiovascular fitness (maximal oxygen consumption [$\dot{V}O_2$ max])[17,18] and high-density lipoprotein cholesterol[19] that are the same as or better than longer questionnaires. The reliability of this question ranges from $r = .60$ to $r = .71$ in women and from $r = .26$ to $r = .36$ in men.[20]

An answer of no to the physical activity question was used to define sedentary behavior; an answer of yes defined nonsedentary behavior. The single activity associated with sweat-related physical activity was categorized into one of 18 types: walking, weight lifting, aerobics/calisthenics, jogging/running, swimming, cycling, gardening, team sports, racquet sports, other aerobics, skiing, dancing, nonaerobic sports, work related, outdoors, farm related, chopping wood, and miscellaneous. Cigarette smoking was evaluated by self-report comparing present vs past smoking or never smoked.

Physical activity was evaluated both as the mean number of times per week participants were involved in "sweat activity" and as the percentage of the population labeled sedentary. Analysis of covariance was used to derive least square means to determine the mean number of times per week of sweat activity, and F tests were

Charles B. Eaton, Anne N. Nafziger, and Thomas A. Pearson are with the Department of Medicine/Epidemiology at Columbia University's Mary Imogene Bassett Research Institute, Cooperstown, NY. David S. Strogatz is with the Department of Epidemiology, State University of New York at Albany and the New York State Health Department's School of Public Health.

Requests for reprints should be sent to Charles B. Eaton, MD, MS, Department of Family Medicine, Brown University/Memorial Hospital of Rhode Island, 111 Brewster St, Pawtucket, RI 02860.

This paper was accepted June 29, 1993.

TABLE 1—Mean Times per Week of Sweat-Related Activity and Percentage of Population Involved in Sedentary Life-Style in the Otsego County Health Census

	Mean Times per Week of Sweat-Related Activity		Sedentary, %		Nonsedentary, %		Total No.	
	Men	Women	Men	Women	Men	Women	Men	Women
Race								
White	2.68	2.34	44.7	47.4	55.3	52.6	12 771	15 021
Black	2.68	2.42	32.8	45.5	67.2	54.5	67	66
Other	2.63	2.73	40.5	37.0	59.5	63.0	74	81
Marital status**								
Married	2.49	2.28	47.5	46.4	52.5	53.6	8 976	9 312
Single	3.25	2.58	33.0	38.1	67.0	61.9	2 630	2 377
Divorced	2.81	3.26	42.8	45.3	57.2	54.7	878	1 383
Widowed	2.28	2.15	61.9	64.5	38.1	35.5	410	2 074
Age**								
<18 y	3.44	2.69	31.4	38.8	68.6	61.2	225	307
18–35 y	3.04	2.45	33.4	38.1	66.6	61.9	3 829	4 360
36–65 y	2.43	2.35	45.7	43.0	54.3	57.0	5 525	6 152
>65 y	2.52	2.09	49.7	56.3	50.3	43.7	1 904	2 617
Location								
Town	2.71	2.34	38.1	43.1	61.9	56.9	4 335	5 581
Rural	2.65	2.34	44.4	44.5	55.6	55.5	7 149	7 575
Level of education*								
<12 y	2.49	2.17	55.8	60.6	44.2	39.4	2 707	2 553
12 y	2.66	2.31	47.0	49.4	53.0	50.6	4 816	5 625
13–16 y	2.79	2.42	36.2	41.2	63.8	58.8	3 986	5 486
>16 y	2.70	2.41	33.3	36.4	66.7	63.6	1 523	155
Occupation**								
Homemaker	1.86	2.20	54.7	54.1	45.3	45.9	64	4 128
Blue collar	2.58	2.06	49.6	53.4	50.4	46.6	3 927	910
Student	3.71	2.95	19.7	31.2	80.3	68.8	787	911
White collar	2.44	2.30	47.3	47.8	52.7	52.2	3 107	5 197
Professional	2.53	2.39	39.7	40.5	60.3	59.5	3 483	3 640
Farm related	3.40	3.08	45.9	46.7	54.1	53.3	1 300	272
Total***	2.57	2.21	44.7	47.5	55.3	52.5	13 908	15 396

*F test $P < .01$ for difference in groups for each gender.
**F test $P < .001$ for difference in groups for each gender.
***F test $P < .001$ for difference in groups.

used to determine differences in activity by race, gender, age, marital status, location, level of education, and occupation. A logistic model comparing sedentary with nonsedentary activity was evaluated separately for men and women stratified by age group (young = 17 to 35, middle age = 36 to 65, old = 65 to 89) using dichotomous predictor variables of smoking, having a blue-collar job, being married, and having less than a high school education, with the quetelet index (kg/m^2) evaluated as a continuous variable. All analyses were done using the SAS statistical package.

Results

Levels of Physical Activity

Table 1 categorizes subjects as sedentary and nonsedentary and describes the mean number of times per week participants engage in sweat activity for each sociodemographic descriptor. In this study, 46.2% were sedentary and 53.8% were nonsedentary. Excluding work and examining only leisure-time physical activity reveals that 48.8% were sedentary.

Comparisons of the mean number of times per week participants were involved in sweat activity revealed differences by age, marital status, gender, education (in all cases, $P < .001$), and occupation ($P < .01$) but no differences by race or location. Farmers demonstrated an increased amount of sweat activity compared with persons in most other occupations but had the same percentage of sedentary behavior, suggesting a bimodal distribution.

Types of Physical Activity

As described in Table 2, walking was the most popular activity for men and women in all age groups. For those aged 18 to 35 and 36 to 65, other prevalent types of activity for men were work related and jogging while women chose aerobics. In the oldest age group (65 to 89), men preferred gardening, work-related activity, and nonaerobic activity while women preferred gardening and cycling.

Correlates of Physical Activity

Table 3 reveals the odds ratios and 95% confidence intervals of the multivariable logistic regression for men and women comparing sedentary levels of activity with nonsedentary levels. The correlates of sedentary activity varied with age and gender.

Discussion

This study demonstrates that a large proportion (46.2%) of residents of this rural county were sedentary. Levels of sedentary behavior are higher in this study than both the New York State Behavioral Risk Factor Surveillance System[21] estimates in 1986 (28.4%) and 1987 (36.2%) and the national Behavioral Risk Factor Survey[11] estimates in 1982 (36%). However, the fact that our study questionnaire asked about sweat-related physical activity and not about less intense activity may have overestimated sedentary levels.

Demographic factors such as age, gender, educational attainment, and marital status; behavioral factors such as cigarette smoking; and physiologic factors such as quetelet index are predicative of sedentary activity compared with nonsedentary activity in this rural population and are consistent with studies in other settings.[1,22]

The single type of activity reported in this study is difficult to compare against the multiple activities recorded by other national surveys.[5–11] Nevertheless, walking (49.4%) is clearly the most popular activity in this study, as it is in other surveys,[5–11] and jogging, aerobics, and cycling (approximately 6%) are of similar rank order compared with their standing in national surveys.[5–11]

The high level of physical activity found in this rural county may reflect a more active life-style associated with rural living. This appears particularly true for male farmers, of whom 46.4% were involved four or more times per week in sweat-related activity, which contributed

TABLE 2—Percentage of Men and Women in Each Type[a] of Physical Activity in the Otsego County Health Census, by Age Group

Type of Activity	18 to 35 y		36 to 65 y		65 to 89 y	
	Men (n = 2689)	Women (n = 2800)	Men (n = 3164)	Women (n = 3648)	Men (n = 1052)	Women (n = 1251)
Walking	22.3	50.5	38.9	66.9	53.7	71.1
Weight lifting	5.8	0.7	1.2	0.3	0.2	0.0
Aerobics/callisthenics	2.8	12.1	2.6	8.3	2.1	3.4
Jogging/running	11.8	8.7	8.0	2.6	0.2	0.1
Swimming	1.6	2.4	2.2	1.9	1.6	2.7
Cycling	7.0	8.7	5.2	5.7	4.0	5.9
Gardening	2.4	1.8	4.4	3.3	13.5	8.6
Team sports	16.3	4.0	4.3	0.5	0.2	0.1
Racquet sports	1.6	1.5	3.3	1.2	0.9	0.6
Other aerobics	1.2	0.8	1.3	0.6	0.3	0.0
Skiing	0.1	0.1	0.1	0.0	0.0	0.0
Dancing	0.1	1.0	0.2	1.0	1.0	0.9
Nonaerobic (golf)	1.3	0.5	3.3	1.0	7.0	2.6
Work related	19.3	6.1	17.8	5.0	8.2	3.4
Outdoors (hunting)	0.8	0.2	0.7	0.0	1.0	0.2
Farm related	4.3	0.8	4.9	1.3	3.3	0.2
Chopping wood	1.1	0.0	1.5	0.0	2.7	0.2
Miscellaneous	0.2	0.2	0.2	0.3	0.3	0.1

[a]Only single categories are recorded.

TABLE 3—Odds Ratios of Correlates of Sedentary Activity in the Otsego County Health Census

	Men		Women	
	Odds Ratio	95% CI[a]	Odds Ratio	95% CI[a]
Ages 18–35 y				
Smoking	1.60	1.33, 1.92	1.46	1.28, 1.73
Quetelet index (kg/m^2)	1.03	1.01, 1.05	1.00	0.99, 1.02
Blue collar	1.32	1.05, 1.66	1.32	0.82, 2.12
Married	0.64	0.54, 0.25	1.45	1.27, 1.66
Less than high school education	1.19	0.92, 1.54	1.02	0.79, 1.30
Ages 36–65 y				
Smoking	1.63	1.41, 1.88	1.45	1.28, 1.65
Quetelet index (kg/m^2)	1.05	1.03, 1.07	1.03	1.02, 1.04
Blue collar	1.22	1.00, 1.49	1.50	1.01, 2.21
Married	0.96	0.78, 1.19	1.19	1.01, 1.40
Less than high school education	1.49	1.23, 1.79	1.34	1.12, 1.59
Ages >65 y				
Smoking	1.61	1.20, 2.15	1.41	1.11, 1.79
Quetelet index (kg/m^2)	1.00	0.97, 1.03	1.01	0.99, 1.03
Blue collar	1.18	0.85, 1.64	1.27	0.69, 2.33
Married	0.81	0.62, 1.05	0.82	0.69, 0.96
Less than high school education	1.53	1.22, 1.91	1.48	1.23, 1.78

[a]CI = confidence interval, estimated by logistic regression.

to an average of 1 more sweat day of activity per week in farming compared with other occupations.

The high frequency of walking and the gender differences in both levels and choice of activity suggest that further research and public policy recommendations focus on these issues. The health impacts of the bimodal distribution of farmers' activity patterns with a large percentage involved in sedentary behavior (43%) and a high percentage (46.4%) involved four or more times per week in sweat activity should be explored in future studies. □

Acknowledgments

Partial support for this research was supplied by the Public Health Service, National Service Research Award 5T32HL07642-04.

The authors gratefully acknowledge Benjamin Rybicki for his assistance in the early statistical analysis and Irene Reis for her assistance in the preparation of the manuscript.

References

1. White CC, Powell KE, Hogelin GC, Gentry EM, Forman MR. The behavioral risk factor surveys: IV. the descriptive epidemiology of exercise. *Am J Prev Med.* 1987; 3:304–310.
2. Montoye JH. *Physical Activity and Health: An Epidemiologic Study of an Entire Community.* Englewood Cliffs, NJ: Prentice Hall; 1975:13–27.
3. Folsom AR, Caspersen CJ, Taylor HL, et al. Leisure time physical activity and its relationship to coronary risk factors in a population-based sample: the Minnesota Heart Survey. *Am J Epidemiol.* 1985;121: 570–579.
4. Dannenberg AL, Keller JB, Wilson PF, Castelli WP. Leisure time physical activity in the Framingham Offspring Study: description, seasonal variation, and risk factor correlates. *Am J Epidemiol.* 1989;129: 76–88.
5. Stephens T, Jacobs DR, White CC. A descriptive epidemiology of leisure-time physical activity. *Public Health Rep.* 1985; 100:147–158.
6. President's Council on Physical Fitness and Sports. National adult physical fitness survey. *Phys Fitness Res Digest.* 1974;4: 1–27.
7. Exercise and participation in sports among persons 20 years of age and over, United States, 1975. *Adv Data Vital Health Stat.* 1978; no. 19.
8. *The Perrier Study: Fitness in America.* New York, NY: Perrier-Great Waters of France, Inc; 1979.
9. Highlights from wave I of the National Survey of Personal Health Practices and Consequences: United States, 1979. *Vital Health Stat [15].* 1981;1. DHHS publication PHS 81–1162.
10. *Canada Fitness Survey: Fitness and Lifestyle in Canada.* Ottawa, Canada: Canada Survey; May 1983.
11. *The Miller Lite Report on American Attitudes toward Sports.* Milwaukee, Wis: Miller Brewing Co; 1983.
12. Enterline PE, Rikli AE, Sauer H, Hyman M. Death rates for coronary heart disease in metropolitan and other areas. *Public Health Rep.* 1960;75:759–766.
13. Miller MK, Stokes CS, Clifford WB. A comparison of rural-urban mortality differential for deaths from all causes, cardiovascular disease and cancer. *J Rural Health.* 1987;3:23–34.
14. Ingram DD, Gillum RF. Regional and urbanization differentials in coronary heart disease mortality in the United States, 1968–1985. *J Clin Epidemiol.* 1989;42:857–868.
15. Najem GR, Hutcheon DE, Feuerman M. Changing patterns of ischemic heart disease mortality in New Jersey 1968–1982, and the relationship with urbanization. *Int J Epidemiol.* 1990;19:26–31.
16. Paffenbarger RS, Wing AL, Hyde RT. Physical activity and index of heart attack risk in college alumni. *Am J Epidemiol.* 1978;108:161–175.
17. Siconolfi SF, Lasater TM, Snow RC, Carleton RA. Self-reported physical activ-

ity compared with maximal oxygen uptake. *Am J Epidemiol.* 1985;122:101–105.
18. Kohl HW, Blair SN, Paffenbarger RS, Macera CA, Kronenfeld JJ. A mail survey of physical activity habits as related to measured physical fitness. *Am J Epidemiol.* 1988;127:1228–1239.
19. Washburn RA, Goldfield SR, Smith KW, McKinlay JB. The validity of self-reported exercise-induced sweating as a measure of physical activity. *Am J Epidemiol.* 1990; 132:107–113.
20. Nafziger AN, Eaton CB, Jenkins PL, Bolin SJ, Pearson TA. Marked gender differences in the reliability and validity of self-reported physical activity ascertained by mail surveys, physician interview and phone survey. *Am J Epidemiol.* 1992;136:992.
21. Bzduch H. Physical activity and exercise. *New York State Behavioral Risk Factor Surveillance System 1986 and 1987.* Albany, NY: New York State Department of Health; January 1990;23:59–67. New York State Department of Health Monographs.
22. Dishman RK, Sallis JR, Orenstein DR. The determinants of physical activity and exercise. *Public Health Rep.* 1985;100:158–170.

New Report from APHA: Access to Care for HIV Disease

The American Public Health Association's Special Initiative on AIDS has recently published the tenth report in its series on acquired immunodeficiency syndrome (AIDS), prepared under the auspices of the APHA AIDS Working Group. The new report is entitled "Access to Care for HIV Disease."

This report summarizes the current knowledge about the manifestations of the disease and the types of services needed by people with HIV. Data on the costs of HIV disease are summarized. Barriers to care are reviewed and strategies for improving access to care are examined.

To date, the series includes these reports:

Report 1: Casual Contact and the Risk of HIV Infection, 2nd ed.
Report 2: Contact Tracing and Partner Notification
Report 3: Illicit Drug Use and HIV Infection, 2nd ed.
Report 4: HIV Antibody Testing
Report 5: Public Health Implications of PCP Prophylaxis
Report 6: Pediatric HIV Infection
Report 7: Public Health Implications of Early Intervention in HIV Disease
Report 8: Women and HIV Disease
Report 9: Tuberculosis and HIV Disease
Report 10: Access to Care for HIV Disease

Each report is $3.50 per copy for APHA members, $5.00 for nonmembers. The complete ten-report series is $45.00 for APHA members, $31.50 for nonmembers. Orders must be prepaid. Order from the American Public Health Association, Publication Sales, Department 5037, Washington, DC 20061-5037; tel. (202) 789-5667.

Car Mass and Fatality Risk: Has the Relationship Changed?

ABSTRACT

Objectives. The finding that the relative safety disadvantage of small compared with large cars is less for post-1980 cars than for pre-1980 cars has stimulated speculation that increasing fuel economy standards would increase fatalities less than previously expected. Fatal crashes between two cars of similar model year were examined to see whether this would be the case.

Methods. Driver fatality risk in relation to car mass was examined with Fatal Accident Reporting System data for crashes between two cars of a specific model year.

Results. The relative risk for driver fatality in the lighter car compared with the other driver's risk in a car 50% heavier was as follows: for 1966 through 1979 cars, the risk was between 3.7 and 5.1; for 1984 cars, 2.6; and for 1990 cars, 4.1.

Conclusions. The results suggest that the lesser mass effect observed for mid-1980s cars occurred because improved crashworthiness features appeared in small cars earlier than in large cars. As all cars are redesigned, the relationship between risk and mass can be expected to approach that observed earlier in pre-1980 cars. If so, future fatality increases from fuel economy increases will be greater than estimated on the basis of mid-1980 data. (*Am J Public Health.* 1994;84:33-36)

Leonard Evans, DPhil, and Michael C. Frick, MA

Introduction

It has been established for decades that when cars of unequal mass crash into each other, driver fatality risk in the lighter car exceeds that in the heavier car.[1] Recent studies have found that the difference in fatality risk for the same mass difference is greater for older cars (model year 1979 or earlier) than for more recent cars (1980 or later).[2-4] Such findings have contributed to suggestions that car mass is exercising a declining effect on fatality risk and that car mass effects on safety might become less important or even disappear in the future. Such suggestions imply that increasing fuel economy standards would lead to a smaller increase in occupant fatalities than would be expected on the basis of current or earlier data.

The purpose of the present study was to determine whether the relationship between driver fatality risk and car mass has changed and to identify the relationship most likely to apply in the future. We used data from crashes between two cars of unequal mass but similar model year to determine how the ratio of drivers killed in lighter cars to those killed in heavier cars depends on the ratio of the masses of the cars.

Methods

The Fatal Accident Reporting System is a computerized data file maintained by the National Highway Traffic Safety Administration that contains detailed information on all fatal traffic crashes in the United States since January 1, 1975.[5] We used data for two-car crashes between 1975 and 1991 in which at least one of the two drivers was killed and for which the mass and model year of each car were coded.

The method was identical to that used in our previous study,[2] except that the analysis was confined to crashes between cars of specific model year. Ideally, we would have liked to focus exclusively on crashes between cars of the same model year; however, there are generally insufficient data to allow such analyses. Instead, we used a moving average approach that includes cars of the previous and subsequent model years. For example, we include any crash between a 1974, 1975, or 1976 model year car and any other 1974, 1975, or 1976 model year car, thereby increasing the expected number of crashes by a factor of nine. The central value of the moving average—in this example, model year 1975—is the independent variable in the analyses.

For each car crash, a mass ratio, μ', was defined as

$$(1) \qquad \mu' = \frac{\text{Mass of Heavier Car}}{\text{Mass of Lighter Car}}.$$

All crashes for each model year were divided into five categories of μ'; as far as possible, the five categories contained equal numbers of crashes. For each category, a driver fatality ratio, R, was defined as

$$(2) \quad R = \frac{\text{No. Driver Fatalities in Lighter Cars}}{\text{No. Driver Fatalities in Heavier Cars}}.$$

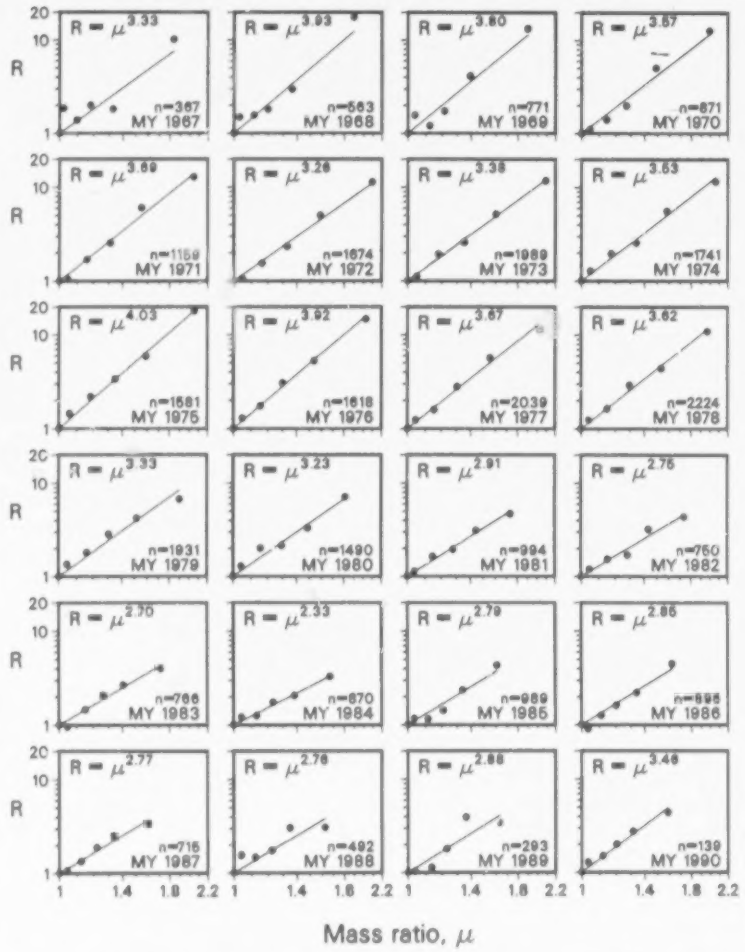
The data were least squares fitted to

$$(3) \qquad \log(R) = u \log(\mu),$$

The authors are with the Operating Sciences Department, General Motors Research, Warren, Mich.

Reprint requests should be sent to Leonard Evans, DPhil, General Motors NAO R&D Center, Automotive Safety and Health Research Department, 30500 Mound Rd, Warren, MI 48090-9055.

This paper was accepted May 18, 1993.



Note. The graph for model year (MY) 1967 includes all crashes in which a 1966, 1967, or 1968 car crashed into a 1966, 1967, or 1968 car, and so forth. The number of crashes contributing to each plot is indicated by n.

FIGURE 1—Driver fatalities in the lighter car divided by those in the heavier car (R) vs the mass of the heavier car divided by the mass of the lighter car (μ).

where μ is the average of the individual values of μ' in the given category. Each point was weighted in inverse proportion to the variance, σ^2, in $\log(R)$. On the assumption that the fatalities arise from a Poisson process, this variance is given by

$$\sigma^2 = 1/(\text{Fatalities in Lighter Car}) + 1/(\text{Fatalities in Heavier Car}). \quad (4)$$

Because $R \equiv 1$ when $\mu = 1$, the fit involves only one parameter, u; crashes between cars of equal mass are excluded from the analyses. Although the fit is to the log-transformed data, a more convenient form of Equation 3 is

$$R = \mu^u. \quad (5)$$

Results

There are relatively few data for the early model years, because by the time the Fatal Accident Reporting System was begun in 1975, most such cars had already been retired from the road. For the most recent model years there are also relatively few data, because, for example, 1990 cars can contribute only to the data for 1989, 1990, and 1991. It will be another 10 or so years before nearly all crashes involving 1990 cars have occurred.

Because of the moving average approach, the results for specific model years are not independent of the data for adjacent model years. For example, the 1581 crashes for model year 1975 arise from the following specific contributions: (4,4) 225; (4,5) 308; (4,6) 386; (5,5) 137; (5,6) 318; and (6,6) 207 crashes, where (4,4) denotes crashes between model year 1974 cars and model year 1974 cars, and so forth. About 176 crashes (1581/9) are expected between cars of the same model year, and 704 between cars whose model years differed by one. Some of these same crashes also contribute to the estimates for model years 1973, 1974, 1976, and 1977. The error in u, Δu, is one standard error.

Because Equation 5 contains only one parameter, u, the relationship it describes is determined completely by choosing any one illustrative point, such as, say, the one corresponding to one of the cars being 50% heavier than the other ($\mu = 1.5$). This provides a possibly more immediately understandable measure than the essentially equivalent parameter u. For example, for model year 1975 we obtain $R = (1.5)^{4.03} = 5.1$, meaning that if two 1975 cars, one of them weighing 900 kg and the other 1350 kg, crash into each other, the driver of the 900-kg car is 5.1 times as likely to be killed as is the driver of the 1350-kg car. The illustrative mass ratio of 1.5 rather than the value 2 used in earlier work[1,2] was chosen because very few cars made after model year 1980 had as much as twice the mass of other cars of the same model year.

Figure 1 shows R vs μ, as determined by dividing the data into five categories of mass ratio, leading to five plotted points for every model year. The choice of five was based on a compromise between the need to have a sufficient number of crashes in each cell for model years for which there were few data and the need for a sufficient number of points to accurately determine Equation 1. Although a case can be made for letting the number of categories increase for model years with more data, we considered it more important to subject the data for each model year to identical treatment. When there are many data, the estimated values of u and Δu depend hardly at all on the number of categories, in accord with statistical theory. When there are few data and noisy relationships, the parameters are more dependent on the number of categories.

Because the parameters for model years 1989 and 1990, which are based on relatively few data, are central to the study, it is important to establish that they do not depend unduly on the arbitrarily chosen number of analysis categories. Using 3, 4, 5, 6, 7, 8, and 9 categories gives values of u for model year 1989 of 3.04, 3.04, 2.88, 2.76, 2.78, 2.92, and 2.90, re-

spectively; the corresponding values for model year 1990 are 3.17, 3.47, 3.46, 3.26, 3.17, 3.10, and 3.05 (the values used are underlined).

The variation of the parameter u as a function of model year (Figure 2) indicates substantial changes. Of particular note is the strictly monotonic decline from $u = 4.03$ for model year 1975 to $u = 2.33$ for model year 1984. This decline is followed by a rapid rise to $u = 3.46$ by model year 1990. The upward trend indicates that the effect of mass on relative driver fatality risk for future model years may become similar to what it was for earlier model years.

The results above are for all drivers, regardless of safety-belt use. To address whether this fact might introduce confounding effects because safety-belt-use rates have increased, we repeated all analyses, excluding all drivers coded as using safety belts. The results showed no systematic departures from those shown in Figures 1 and 2. However, scatter increased considerably because of smaller sample sizes; the total number of available crashes for model years 1984 through 1990 decreased by more than 50%. Such large reductions occur because if either driver is coded as belted, the crash is excluded. After laws requiring safety-belt use became widespread, nonwearers, especially surviving drivers, were increasingly likely to be incorrectly coded as wearers.[1]

The decision to include drivers regardless of safety-belt use is not expected to have had much effect on the pattern shown in Figure 2. A number of studies using a variety of methodologies have indicated that safety-belt effectiveness does not depend systematically on car mass.[1,2] The study that most closely parallels the present one found a value of u in Equation 5 for unbelted drivers of 3.58, compared with $u = 3.53$ for drivers regardless of belt use and $u = 3.60$ for belted drivers.[2] Even if belt-wearing rates did depend on car mass, this would not affect results much because belt use was relatively low for all model years (27% of fatally injured drivers for model year 1987 were coded as belted, the highest rate for any model year) and belt effectiveness is lower in two-car crashes (30% ± 8%) than in single-car crashes (62% ± 5%).[6] Additional analysis found no important relationships between belt-wearing rates and car mass for any model year.

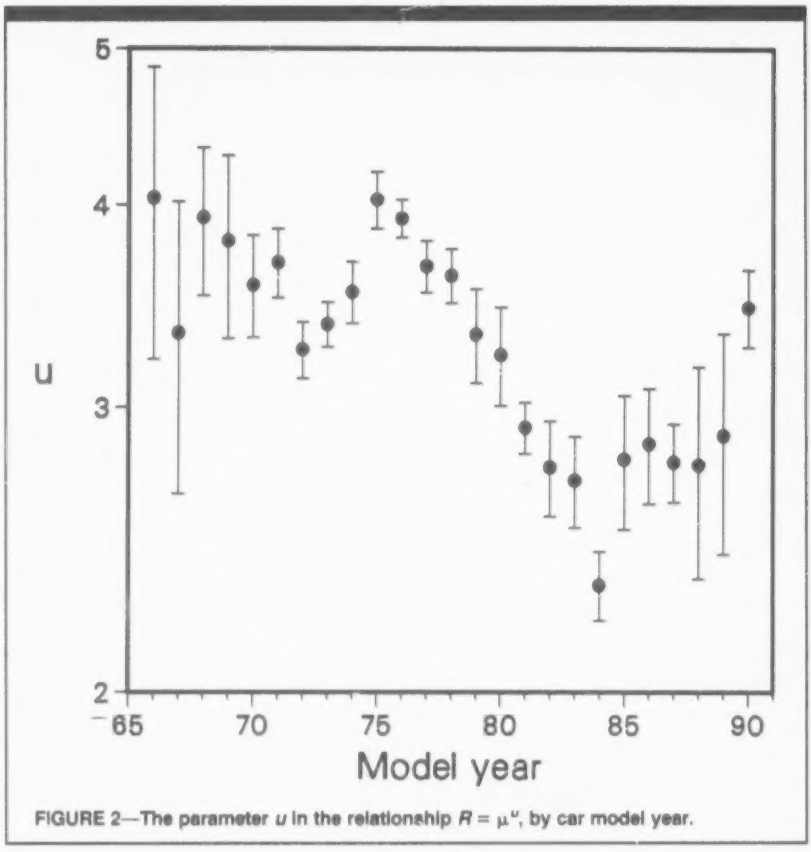


FIGURE 2—The parameter u in the relationship $R = \mu^u$, by car model year.

Discussion

The main purpose of a phenomenological study like the present one is to discover what is happening. Although the results of such studies do not identify the mechanisms leading to the observed effects, they may offer insights into such mechanisms. The following interpretation of Figure 2 is offered as a plausible, though not demonstrated, explanation.

One can interpret a relationship between car mass and fatality risk as including two contributions. First, a contribution is made by what we refer to as the intrinsic effect of mass; included with the intrinsic effect of mass is the effect of size, because these are so strongly correlated.[3] Second, there is a design-related factor. Particular small cars may differ from particular large cars in ways not intrinsically related to their size; for example, if a small car with a rear engine crashed head-on into a large car with a front engine, the relative fatality risk might be larger than if both cars had similarly located engines. Contributions from such differences in some particular early models might generate a small effect on the overall relationships. Apart from this factor, we interpret the car mass effect for pre-1970 cars as reflecting an intrinsic mass effect for a fleet of cars of relatively homogeneous design features but differing in mass. Thus $u \cong 3.7$ represents the mass effect when two cars with similar design features crash into each other.

From the mid-1970s onward, some vehicles were subject to major redesign in which new methods and technologies were applied to manage crash energy dissipation in ways calculated to reduce the injury risk to occupants. Small cars tended to be the first to benefit from such redesign. It thus seems plausible that the lower values of u for mid-1980s cars (for example, $u = 2.3$ for model year 1984) reflect design as well as intrinsic mass effects. Thus the greater risk associated with the smaller car was partially offset by the smaller car's having crashworthiness features not yet incorporated into larger cars. It is also possible that redesign resulted in stiffer structures, which might reduce risks to occupants traveling in the redesigned car but increase risks to occupants in struck cars. This could have a large effect on the fatality ratio, because the numerator would decline while the denominator increased.

As the 1980s proceeded, large cars were subject to the same redesign attention previously devoted to small cars. Thus, as the crashworthiness of the large

cars increased, the relative disparity in risk appeared to be moving back to reflect the same intrinsic mass effect that applied to the earlier model years. At present it is not possible to determine whether values of u will continue to increase and will approach, reach, or exceed their previous levels. However, the trend is consistent with the interpretation that u may approach a new stable value of ~3.7, which represents the intrinsic effect of mass when cars are otherwise of homogeneous design. The much lower u values of the mid-1980s cars were a temporary aberration reflecting effects of both mass and design. Figure 2 shows definitively that the disparity in risk between small and large cars is not continuing to decline and is now clearly greater than the minimum reached for model year 1984.

The results presented here offer a clear interpretation of earlier unexplained findings[2-4] of large differences between older (model year 1979 and earlier) and newer (model year 1980 and later) cars. The u values of 3.53 for older and 2.75 for newer cars[2] are in accord with the values shown in Figure 1 (taking into account weighting by the number of data). Conclusions based on an assumption that the results for model years 1980 and later would apply into the future underestimate the effect of mass on safety. The replacement of heavier cars by lighter cars is expected to result in an increase in fatalities substantially greater than was previously estimated.[4] Indeed, calculations based on the car mass–fatality relationship for older cars may estimate future effects more accurately than those based on the relationship for newer cars.

All results presented in this paper are presented in terms of relative risk; they suggest that in the future, occupant risk in a small car will exceed that in a large car by a factor similar to that observed in the past. However, risks to drivers in all cars are expected to decline in the future, as they have in the past.[1,7] Factors that uniformly reduce crash risk do not affect the relationships reported here, even though they reduce the absolute number of driver fatalities. For example, if the nation were to become teetotal, traffic fatalities in essentially all categories would be reduced by approximately half.[1] As a result, the effectiveness of all occupant protection devices (and other safety measures) in reducing the absolute number of fatalities would likewise be halved, but without affecting relative risk relationships such as those reported here. Universal use of safety belts or airbags might have a small effect on the relative risk ratios reported (in addition to the main effect of reducing casualties in all types of vehicles). Although belts offer similar proportionate risk reduction in cars of all sizes, as discussed earlier, a recent finding[8] suggests a somewhat higher effectiveness in larger cars. As airbags also appear to be more effective in large cars than in small cars,[9] universal use of belts and airbags will tend to further increase the already great relative disadvantage in the smaller car. The absolute risk, rather than the relative risk, is, of course, more relevant in discussions of the cost-effectiveness of such expensive occupant protection features as increased vehicle mass or airbags.

The present study has, of necessity, focused exclusively on two-car crashes. Although about 78% of car occupants who are killed are killed in crashes other than two-car crashes,[1] the general conclusions reported here are likely to apply also to other crash types for which direct examination is not technically possible (assuming the validity of the explanation in terms of vehicle redesign).

Conclusions

When a car of a given model year crashes into another car of the same model year but 50% heavier, the driver in the lighter car is more likely than the other driver to be killed by a factor of about 3.7 to 5.1 for 1966 through 1979 model year cars, 2.6 for 1984 model year cars, and 4.1 for 1990 model year cars. These changes appear to be related to vehicle redesign. The smaller effect of car mass for the mid-1980s model years may be a result of smaller cars' containing crashworthiness improvements not yet introduced into larger cars.

In the future the relative safety disadvantage of smaller cars is expected to more closely resemble the larger effects found for pre-1979 model year cars; recent estimates based on data for all model year 1980 or later cars underestimate the extent to which driver fatality risk increases with decreasing car mass. Accordingly, future fatality increases resulting from increases in corporate average fuel economy standards will likely be greater than estimated on the basis of mid-1980 data. □

References

1. Evans L. *Traffic Safety and the Driver.* New York, NY: Van Nostrand Reinhold; 1991.
2. Evans L, Frick MC. Mass ratio and relative driver fatality risk in two-vehicle crashes. *Accid Anal Prev.* 1993;25:213–224.
3. Evans L, Frick MC. Car size or car mass—which has greater influence on fatality risk? *Am J Public Health.* 1992;82:1009–1112.
4. Evans L, Frick MC. Driver fatality risk in two-car crashes—dependence on masses of driven and striking car. In: *Proceedings of the Thirteenth International Technical Conference on Experimental Safety Vehicles, Paris, France, 4–7 November 1991.* Washington, DC: National Highway Traffic Safety Administration; July 1993:83–93. Document DOT HS 807 990, vol. 1.
5. *Fatal Accident Reporting System 1989.* Washington, DC: National Highway Traffic Safety Administration; March 1991. Document DOT HS 807 693.
6. Evans L, Frick MC. Safety belt effectiveness in preventing driver fatalities versus a number of vehicular, accident, roadway and environmental factors. *J Safety Res.* 1986; 17:143–154.
7. Evans L. How safe were today's older drivers when they were younger? *Am J Epidemiol.* 1993;137:769–775.
8. Evans L. Driver injury and fatality risk in two-car crashes versus mass ratio inferred using Newtonian Mechanics. *Accid Anal Prev.* In press.
9. O'Neill B, Lund AK. The effectiveness of air bags in preventing driver fatalities in the United States. Arlington, Va: Insurance Institute for Highway Safety; October 1992.

Cesarean Section and Operative Vaginal Delivery in Low-Risk Primiparous Women, Western Australia

Anne W. Read, PhD, Walter J. Prendiville, FRACOG, Vivienne P. Dawes, MBChB, and Fiona J. Stanley, MD

ABSTRACT

Objectives. A major component of the increasing trend in cesarean sections in Western Australia is the rise in emergency cesarean sections in primiparous women. The aim of this study was to identify independent risk factors (particularly those known early in pregnancy) associated with operative delivery in low-risk primiparous women.

Methods. Retrospective multivariate logistic regression analyses of antenatal and perinatal data were conducted for all low-risk primiparous women entering labor spontaneously and giving birth in Western Australia in 1987 (n = 3441).

Results. Of the subjects, 58% had a spontaneous vaginal delivery, 8% had an emergency cesarean section, and 34% had an operative vaginal delivery. The significant independent risk factors for emergency cesarean section were older maternal age, shorter maternal height, heavier infant birthweight, and long labor. The risk factors for operative vaginal delivery were older maternal age, shorter maternal height, heavier infant birthweight, epidural anesthesia, labor/delivery complications, male infant, private patient status, and being married.

Conclusions. This multivariate analysis confirms known risk factors for operative delivery in low-risk primiparous women and suggests that it may be possible to predict the likelihood of operative delivery for an individual woman by using knowledge of maternal age and height and assessment of infant birthweight. (*Am J Public Health.* 1994;84:37–42)

Introduction

Many countries have recently experienced a progressive rise in the incidence of cesarean section.[1,2] Western Australia, which witnessed an increase from 5% to 19% in the proportion of women delivered by cesarean section from 1975 to 1990, is included in this trend.[3] Cesarean sections, while usually undertaken in the interests of the offspring, may result in complications for the mother.[4] Accordingly, the wisdom of the increasing trend is in question, and efforts have been made to stabilize or reverse this situation.[5]

In a previous paper,[2] we identified the two major components of the increasing trend in Western Australia as the rise in emergency cesarean sections in primiparous women and the rise in repeated elective cesarean sections in multiparous women. Clearly, these two circumstances are related in that if a first birth is delivered by cesarean section, then subsequent births are more likely to be delivered by the same method. Therefore, it is important to investigate emergency cesarean sections in primiparas with the aim of ascertaining why they are being performed more frequently. During the 1980s, as cesarean sections increased in frequency, operative vaginal deliveries decreased.[2] Thus, we considered it important to investigate all types of operative delivery in the population.

This study was based on the total population of apparently low-risk primiparas who began labor spontaneously at term with a cephalic presentation of a singleton infant. We sought to identify, using logistic regression analysis, the significant risk factors associated with an increased chance of either operative vaginal (forceps or vacuum extraction) or cesarean section delivery in this population.

Methods

This study was carried out on existing computer files of the Western Australian Midwives' Notification of Case Attended Form 2, which are linked to the Western Australian Hospital Morbidity Data System, for the year 1987. The Midwives' Form is a statutory document completed by the midwife in attendance at each live birth and stillbirth in Western Australia, provided the infant is at least of 20 weeks gestation or 400 g birthweight.[3] There are about 25 000 births annually in Western Australia.

All primiparous women with low-risk pregnancies who gave birth in Western Australia in 1987 were selected. We defined low-risk pregnancies as follows: (1) singleton infant, (2) no recorded medical conditions or pregnancy complications, (3) vertex presentation of the infant, and (4) spontaneous commencement of labor at term (at least 37 weeks).

Women whose pregnancies fulfilled these criteria were divided into three groups according to whether they had a spontaneous vaginal, operative vaginal

Anne W. Read is with the Western Australian Research Institute for Child Health, Perth. Walter J. Prendiville is with the University Department of Obstetrics and Gynaecology, King Edward Memorial Hospital for Women, Subiaco, Australia. Vivienne P. Dawes is with the Epidemiology Branch of the Health Department of Western Australia, Perth. Fiona J. Stanley is with the Department of Paediatrics, University of Western Australia, and the Western Australian Research Institute for Child Health.

Correspondence should be sent to Anne W. Read, PhD, Western Australian Research Institute for Child Health, GPO Box D184, Perth 6001, Western Australia.

This paper was accepted with revisions May 3, 1993.

(vacuum or forceps), or emergency cesarean section delivery. There were no elective cesarean sections because these women, by definition, all had spontaneous onset of labor.

Frequencies of selected demographic and perinatal characteristics for women in the three groups were coded, as shown in Table 1, in preparation for multivariate analyses. The "other" category of maternal race comprised mostly (78%) Asian women. With regard to patient classification, the universal health care system in Australia provides public hospital care by hospital doctors. In addition, people may obtain insurance that provides for treatment in private or public hospitals by doctors of their own choice. We chose a priori to investigate the variables in Table 1, together with maternal age and height, either because they have known associations with pregnancy outcome or because they are apparent early in the pregnancy and can therefore be useful predictor variables if shown to be significant. Maternal age and height can always be determined at the first antenatal visit, and we were particularly interested in these variables as potential independent risk factors for operative delivery.

Egret software (Statistics and Epidemiology Research Corporation, Seattle, Wash) was used to perform two separate multiple logistic regression analyses.[6] First, we identified the significant independent risk factors for emergency cesarean section relative to spontaneous vaginal delivery, and, second, we carried out a similar analysis for operative vaginal relative to spontaneous vaginal delivery.

A two-stage logistic regression procedure was used to fit the initial models. In the first stage, the demographic variables (maternal race, maternal area of residence, marital status, patient classification, maternal height, and maternal age) were added, followed by the perinatal variables (infant gender, infant birthweight, hours of labor, labor/delivery complications, and anesthesia/analgesia) in the second stage. At each stage, the nonsignificant variables were left in the model. The anesthesia/analgesia and labor/delivery complications variables were not used in the emergency cesarean section analysis because these items were recorded as positive for almost all women who had emergency cesarean sections (the labor/delivery complications were the indications for cesarean section). Eight women whose maternal height was not known and 44 whose patient classification was not known were excluded from these analyses (a total of 51 women).

In the models described here, the coefficient for each variable was adjusted for the effects of all other variables, however, the nonsignificant variables ($P > .05$) are not shown in the tables (with the exception of "patient classification" in Table 3). Maternal height, maternal age, and infant birthweight were factored to allow comparison with the strata at lowest risk for operative delivery according to previous literature (i.e., maternal height of 165 cm or more, maternal age of less than 20 years, and infant birthweight of less than 3000 g).[7–10]

In addition to the initial models being fitted, separate models were fitted to obtain estimated odds ratios (unadjusted for other risk factors) for emergency cesarean section and operative vaginal delivery for each maternal age/maternal height category. Further models were fitted to obtain unadjusted estimated odds ratios for each maternal age/maternal height/infant birthweight category; these values provided an estimate of the risk involved for a woman of a particular age and height according to the birthweight grouping of her infant.

Results

Of the 3641 primiparas with low-risk pregnancies who gave birth in Western Australia in 1987, 2116 (58%) had a spontaneous vaginal delivery, 1242 (34%) had an operative vaginal delivery (642 forceps, 600 vacuum), and 283 (8%) had an emergency cesarean section.

Table 1 shows the numbers and percentage distribution of women in each delivery category for selected demographic and perinatal variables. Labor or delivery complications were recorded for all but three of the women who had emergency cesarean sections. The indication for cesarean section for two of these women was long labor; the third gave birth to an infant who required intubation. Women could have had more than one complication (indication) recorded, and those recorded most frequently were cephalopelvic disproportion (56%), fetal distress (40%), and failure to progress (27%). In addition, 16% had other complications recorded, such as antepartum hemorrhage and prolapsed cord. By contrast, labor and delivery complications were recorded for only 48% of the women who had operative vaginal deliveries (3% had cephalopelvic disproportion, 20% had fetal distress, 10% failed to progress, and 21% experienced other complications). However, for some of the women in this group, the processes that led to an operative delivery would have been recorded elsewhere (e.g., a long labor would have been recorded under "hours of labor").

Women who were younger and taller were less likely to have had operative deliveries than women who were older and shorter (Table 2). Generally, within each maternal height category, the percentages of women with operative deliveries increased with increasing maternal age. Within each maternal age category, the percentages of women with emergency cesarean sections increased with decreasing height; however, this trend was not seen for operative vaginal deliveries, except in those women 35 years of age or more (Table 2).

Table 3 shows the results of the multiple logistic regression analysis in which significant independent risk factors for emergency cesarean section were compared with those for spontaneous vaginal delivery. Among the potential demographic risk factors analyzed, the most important were maternal height and maternal age, the odds ratio increasing with decreasing maternal height and increasing maternal age. All of the selected variables, with the exception of anesthesia/analgesia and labor/delivery complications, were included in this model; however, the area of residence, marital status, maternal race, and infant gender variables were not significant when all of the other variables were entered (data not shown). Although the patient classification variable was not significant in the model, it has been included in Table 3 because it represents an area of special interest with regard to the debate concerning the progressive rise in cesarean births. Private patient status was significantly associated with the outcome of emergency cesarean section until maternal age was added to the model. There were few private patients in the younger age groups; thus, any association between being a private patient and having an emergency cesarean section was due to private patients being older (data not shown). With regard to the perinatal variables, heavier infant birthweight and long labor were significant independent risk factors for emergency cesarean section.

Short maternal height was a significant risk factor for operative vaginal delivery, although there was little difference in the odds ratios for women less than 160 cm in height and those 160 to 164 cm in height. Maternal age was also a significant

predictor, the odds ratio for operative vaginal delivery increasing with each older group (Table 4). In contrast to the risks for emergency cesarean section, private patient classification and being married remained significant risk factors for operative vaginal delivery, even when the other variables were added to the model. As with the emergency cesarean section analysis, infant birthweight was a significant independent risk factor for operative vaginal delivery. Both labor/delivery complications and anesthesia/analgesia were significant independent risk factors for operative vaginal delivery when all other variables were included in the model, and epidural anesthesia was the most significant risk factor of all those included in the model (Table 4).

Table 5 shows that the estimated odds ratios increased as maternal age increased and maternal height decreased; thus, women 35 years of age or more and less than 160 cm tall were at the highest risk. In Table 6, the estimated odds ratios increased with maternal age within each height category but did not vary greatly between the height groups.

Estimated odds ratios for each maternal age/maternal height/infant birthweight category (unadjusted for other factors) were also obtained for both emergency cesarean section and operative vaginal delivery. The complete results from this analysis are not shown because there were small numbers in some of the cells. For example, in Table 5, there were fewer than five women in three of the categories, and these women would have been subdivided further into the three birthweight categories. However, with larger numbers of deliveries available for analysis, it would be useful to examine the combined risk for operative delivery according to birthweight of the infant as well as age and height of the mother. As an example, in one of our categories with a satisfactory number of women for analysis (those 25 to 29 years of age and 160 to 164 cm tall), the estimated odds ratios for emergency cesarean section relative to spontaneous vaginal delivery were 7.2 if the infant's birthweight was less than 3000 g, 14.9 if birthweight was 3000 through 3499 g, and 34.0 if birthweight was 3500 g or more. The estimated odds ratios for operative vaginal relative to spontaneous vaginal delivery for women in the same age/height category were 3.3 if the infant's birthweight was less than 3000 g, 5.4 if birthweight was 3000 through 3499 g, and 9.4 if birthweight was 3500 g or more.

TABLE 1—Maternal and Infant Characteristics of Primiparous Women with Low-Risk Pregnancies, by Method of Delivery: Western Australia, 1987

Characteristic (Regression Analysis Code)	Method of Delivery: Spontaneous Vaginal (n = 2116) No.	%	Operative Vaginal (Vacuum n = 600, Forceps n = 642) No.	%	Emergency Cesarean Section (n = 283) No.	%
Maternal						
Marital status						
Unmarried (1)	468	22.1	152	12.2	44	15.8
Married[a] (2)	1648	77.9	1090	87.8	239	84.5
Race						
Caucasian (1)	1917	90.6	1128	90.8	247	87.3
Aboriginal (2)	93	4.4	28	2.3	10	3.5
Other (3)	106	5.0	86	6.9	26	9.2
Area of residence						
Rural (1)	737	34.8	326	26.3	82	29.0
Metropolitan (2)	1379	65.2	916	73.8	201	71.0
Patient classification						
Public (1)	1160	55.0	463	37.3	132	46.6
Private (2)	910	43.0	778	62.6	151	53.4
Unknown (9)	43	2.0	1	0.1	0	0
Length of labor, h						
<13 (1)	1871	88.4	990	79.7	192	67.8
13+ (2)	245	11.6	252	20.3	91	32.2
Anesthesia/analgesia						
None/other[b] (1)	1957	92.5	706	56.8	0	0
Epidural (2)	148	7.0	527	42.4	176	62.2
General (3)	11	0.5	9	0.7	107	37.8
Labor/delivery complications						
No (1)	1531	72.4	648	52.2	3	1.1
Yes (2)	585	27.6	594	47.8	280	98.9
Infant						
Birthweight, g						
<3000 (1)	463	21.9	147	11.8	29	10.3
3000–3499 (2)	980	46.3	519	41.8	115	40.6
3500+ (3)	673	31.8	576	46.4	139	49.1
Gender						
Female (1)	1095	51.8	547	44.0	119	42.1
Male (2)	1021	48.3	695	56.0	164	58.0

[a]Includes stable de facto relationships.
[b]Includes inhalants (e.g., nitrous oxide), narcotics (e.g., pethidine), and caudal/pudendal nerve blocks.

Discussion

Most studies of obstetric databases have been descriptive. More recently, multivariate techniques have been described in order to control for confounding factors so that significant independent risk factors for obstetric outcomes may be ascertained and quantified.[11] However, there have been few population-based studies, such as ours (which included all low-risk primiparas giving birth in Western Australia in 1987), that use multivariate analyses and include factors such as maternal age and height and infant birthweight.

The operative delivery rates in this low-risk group of women were found to be high. While similar proportions may be found in other units and populations of the Western world, lower operative delivery rates in low-risk primiparas have been reported by individual hospitals in Ireland (cesarean section rate = 6%, operative vaginal rate = 19%) and the United States (cesarean section rate = 12%, operative vaginal rate = 27%).[12] However, a study of a comparable Australian population

TABLE 2—Distribution of Method of Delivery in Each Maternal Age/Maternal Height Category for Primiparous Women with Low-Risk Pregnancies: Western Australia, 1987

Maternal Height, cm[a]	Maternal Age, y	Method of Delivery: Spontaneous Vaginal No.	Spontaneous Vaginal %[b]	Operative Vaginal No.	Operative Vaginal %[b]	Emergency Cesarean Section No.	Emergency Cesarean Section %[b]	Total
<160	<20	116	73.9	27	17.2	14	8.9	157
	20–24	190	57.4	104	31.4	37	11.2	331
	25–29	168	47.6	132	37.4	53	15.0	353
	30–34	47	41.6	47	41.6	19	16.8	113
	35+	5	19.2	13	50.0	8	30.8	26
	Total	526	...	323	...	131	...	980
160–164	<20	101	69.7	40	27.6	4	2.8	145
	20–24	229	60.6	120	31.7	29	7.7	378
	25–29	220	50.8	175	40.4	38	8.8	433
	30–34	77	53.8	55	38.5	11	7.7	143
	35+	13	39.4	16	48.5	4	12.1	33
	Total	640	...	406	...	86	...	1132
165+	<20	120	78.4	30	19.6	3	2.0	153
	20–24	329	65.4	160	31.8	14	2.8	503
	25–29	367	61.0	207	34.3	28	4.7	602
	30–34	105	50.2	92	44.0	12	5.7	209
	35+	24	44.4	22	40.7	8	14.8	54
	Total	945	...	511	...	65	...	1521

[a]Eight women whose heights were not known were excluded.
[b]Percentage of all primiparous women with low-risk pregnancies.

TABLE 3—Logistic Regression Analysis of Risks for Emergency Cesarean Section: Primiparous Women with Low-Risk Pregnancies, Western Australia, 1987

Risk Factor	OR (95% CI)	P
Private patient	1.32 (0.97, 1.78)	.074
Maternal height, cm[a]		
160–164	2.43 (1.70, 3.48)	<.001
<160	5.30 (3.72, 7.56)	<.001
Maternal age, y[b]		
20–24	1.96 (1.12, 3.42)	.019
25–29	2.85 (1.60, 5.06)	<.001
30–34	3.31 (1.73, 6.35)	<.001
35+	11.87 (5.24, 26.88)	<.001
Infant birthweight, g[c]		
3000–3499	1.92 (1.24, 2.99)	.004
3500+	4.16 (2.66, 6.51)	<.001
Long labor (≥ 13 h)	3.74 (2.74, 5.09)	<.001

Note. The nonsignificant variables of area of residence, marital status, maternal race, and infant gender are not shown, although they were included in the model. OR = odds ratio; CI = confidence interval.
[a]Odds ratios relative to maternal height of 165 cm or more.
[b]Odds ratios relative to maternal age of less than 20 years.
[c]Odds ratios relative to birthweight of less than 3000 g.

giving birth in Sydney teaching hospitals found a cesarean section rate of 5% and a forceps delivery rate of 43%.[13] A degree of caution is needed when comparing different publications referring to low-risk populations, even when relatively strict inclusion criteria, such as those we have described here, are used. For example, including only primiparas in spontaneous labor excludes those whose labors are induced; indications for induction of labor vary considerably from one hospital to another and will affect the population "allowed" to proceed to spontaneous labor.

In our multivariate model, private patients were not significantly more likely to have an emergency cesarean section than were public patients, but they were more likely to have an operative vaginal delivery. Cary[14] studied 24- to 25-year-old low-risk primiparas in Queensland, Australia, and found more cesarean section (15% vs 10%) and operative vaginal deliveries (32% vs 21%) in private than public patients. However, Cary did not control for maternal height, infant birthweight, or other potential confounding factors. Hewson et al.[13] studied a similar group of low-risk primiparas in Sydney teaching hospitals and found no significant differences between private and public patients with regard to mode of delivery. In a study of low-risk women in the United States, Haynes de Regt et al.[15] found significantly higher proportions of cesarean sections in private than public patients after allowing for age and parity in one analysis and infant birthweight category in another. In a low-risk population in the United Kingdom, Klein et al.[16] found a significantly higher proportion of forceps deliveries in private than public patients, but there was no significant difference in these two groups with respect to cesarean sections. This was a result similar to that of our study, but Klein et al. did not control for potential confounding factors.

The strong independent effect of maternal age was an important confounding factor when investigating the public/private patient dichotomy with regard to emergency cesarean section, and maternal age may have more influence in primiparas than multiparas.[2,9] Although primiparas of advanced maternal age (e.g., over 35 years) have long been thought to be more likely to have an operative delivery,[7] it is apparent from these analyses that primiparous women 20 to 30 years of age are also more likely to have such an outcome than are primiparas less than 20 years old. The precise explanation for this age-related trend remains uncertain.

Yudkin and Redman[17] found dystocia to occur more frequently in older primiparas than in teenagers. It may be that teenagers are physically more fit for the processes of labor and delivery than are older women. A recent Israeli study using multivariate analysis found a decrease in the risk of cesarean section for teenagers and an increase for women 37 years of age or more, the authors noting that the management of labor may be different in teenagers than in older women.[11] Also, Gordon et al.[18] found that maternal age per se may influence an accoucheur's decision regarding method of delivery, and this may have been the case in our population.

Short maternal height predisposes women to operative delivery.[7,8] Our study has documented an increased chance of operative delivery for low-risk primiparas of both short (less than 160 cm) and medium stature (160 to 164 cm) relative to taller women. Moreover, these increases were independent of other potential and real risk factors such as maternal age, race, and marital status and infant birth-

TABLE 4—Logistic Regression Analysis of Risks for Operative Vaginal Delivery: Primiparous Women with Low-Risk Pregnancies, Western Australia, 1987

Risk Factor	OR (95% CI)	P
Private patient	1.74 (1.45, 2.10)	<.001
Married	1.46 (1.12, 1.91)	.006
Maternal height, cm[a]		
160–164	1.38 (1.14, 1.68)	<.001
<160	1.32 (1.07, 1.64)	.011
Maternal age, y[b]		
20–24	1.44 (1.05, 1.99)	.025
25–29	1.59 (1.14, 2.21)	.006
30–34	2.03 (1.39, 2.97)	<.001
35+	2.94 (1.68, 5.15)	<.001
Male infant	1.33 (1.13, 1.57)	<.001
Infant birthweight, g[c]		
3000–3499	1.52 (1.19, 1.95)	<.001
3500+	2.52 (1.96, 3.24)	
Labor complications	2.44 (2.06, 2.90)	<.001
Epidural	8.37 (6.73, 10.41)	<.001

Note. The nonsignificant variables of area of residence, maternal race, and long labor are not shown, although they were included in the model. OR = odds ratio; CI = confidence interval.
[a]Odds ratios relative to maternal height of 165 cm or more.
[b]Odds ratios relative to maternal age of less than 20 years.
[c]Odds ratios relative to birthweight of less than 3000 g.

TABLE 5—Emergency Cesarean Section Odds Ratios for Each Maternal Age/Maternal Height Category: Primiparous Women with Low-Risk Pregnancies, Western Australia, 1987

	Maternal Height, cm					
	<160		160–164		165+	
Maternal Age, y	No.[a]	Odds Ratio	No.[a]	Odds Ratio	No.[a]	Odds Ratio
<20	14	4.1	4	2.1	3	1.0
20–24	37	8.1	29	4.0	14	1.9
25–29	53	12.6	38	6.3	28	3.0
30–34	19	14.7	11	7.3	12	3.5
35+	8	44.1	4	22.1	8	10.7

Note. One woman whose height was not known was excluded.
[a]Number of women with emergency cesarean sections.

TABLE 6—Operative Vaginal Delivery Odds Ratios for Each Maternal Age/Maternal Height Category: Primiparous Women with Low-Risk Pregnancies, Western Australia, 1987

	Maternal Height, cm					
	<160		160–164		165+	
Maternal Age, y	No.[a]	Odds Ratio	No.[a]	Odds Ratio	No.[a]	Odds Ratio
<20	27	1.2	40	1.2	30	1.0
20–24	104	2.2	120	2.2	160	1.8
25–29	132	3.0	175	2.9	207	2.4
30–34	47	3.7	55	3.7	92	3.0
35+	13	5.3	16	5.3	22	4.4

Note. Two women whose heights were not known were excluded.
[a]Number of women with operative vaginal delivery.

weight. For primiparas, the increasing trend in emergency cesarean section rates with decreasing maternal height has also been shown by Yudkin and Redman.[17]

As illustrated in Tables 5 and 6, maternal age and height can be used together to provide an indication of the increased risk for operative delivery in this population as maternal age increases and maternal height decreases. At every age, women of medium height were twice as likely as taller women to have a cesarean section, and women of short stature were four times more likely to undergo the procedure. At every height, the likelihood of a cesarean section was 10 times greater for women 35 years of age or more than for teenagers. In contrast, height was not as important for operative vaginal delivery, but age was strongly related.

Turner et al.,[10] when investigating nulliparas, found that heavier infant birthweight was a significant risk factor for both cesarean section and forceps delivery for dystocia. In our study, the relationship between birthweight and operative delivery persisted when adjusting for confounding factors such as maternal age and height and infant gender.

Because the timing of epidural anesthesia during the course of labor was not available in our database, we did not investigate this procedure with regard to cesarean section delivery. However, because of the likelihood that most epidurals would have been instituted in early labor and the importance of epidural anesthesia in the progress of labor and type of delivery,[19–21] we included it in our regression models for operative vaginal delivery. Here we found it to be the most important independent risk factor, the relative odds being approximately three times greater than those for labor and delivery complications. Several studies,[13,19,20] including randomized controlled trials,[21] have shown that epidural anesthesia is likely to lead to longer labor and operative delivery.

Our analysis has shown the value of using multivariate models to identify risk factors for operative delivery, particularly when comparing outcomes for different hospitals. Investigations of modifiable risk factors and evaluations of different forms of management of labor should use similar techniques. □

Acknowledgments

The TVW Telethon Foundation (Western Australia) and the Wellcome Trust (United Kingdom) provided financial support.

Portions of this paper were presented at the International Conference on Primary Care Obstetrics and Perinatal Health, 's Hertogenbosch, the Netherlands, March 1991.

We wish to acknowledge the staff at the Health Department of Western Australia, particularly Vivien Gee, for the provision of midwives' data. We thank Maxine Croft, John Gibbins, and Carol Garfield, who maintained the Maternal and Child Health Research Data Base at the Western Australian Research Institute for Child Health, and Dr Nicholas de Klerk for statistical advice.

References

1. Notzon FC. International differences in the use of obstetric interventions. *JAMA.* 1990;263:3286–3291.
2. Read AW, Waddell VP, Prendiville WJ, Stanley FJ. Trends in caesarean section in Western Australia, 1980–87. *Med J Aust.* 1990;153:318–323.

3. Gee V. *Perinatal Statistics in Western Australia. Eighth Annual Report of the Western Australian Midwives' Notification System 1990.* Perth, Western Australia: Health Department of Western Australia; 1991.
4. Petitti D. Maternal mortality and morbidity in cesarean section. *Clin Obstet Gynecol.* 1985;28:763–769.
5. Stafford RS. Alternative strategies for controlling rising cesarean section rates. *JAMA.* 1990;263:683–687.
6. Breslow NE, Day NE. *Statistical Methods in Cancer Research; Vol. I: The Analysis of Case-Control Studies.* Lyon, France: International Agency for Cancer Research; 1980. IARC scientific publication 32.
7. Chamberlain G, Elliot P, Howlett B, Masters K. *British Births 1970; Vol. 2: Obstetric Care.* London, England: Heinemann Medical Books; 1978.
8. Mahmood TA, Campbell DM, Wilson AW. Maternal height, shoe size, and outcome of labour in white primigravidas: a prospective anthropometric study. *Br Med J.* 1988;297:515–517.
9. Holst K, Hilden J, Henningsen I. The unborn and newborn child. 1: Risk factors predicting complicated delivery in a general population of 4,102 women. *Acta Obstet Gynecol Scand.* 1989;68:699–706.
10. Turner MJ, Rasmussen MJ, Boylan PC, Macdonald D, Stronge JM. The influence of birth weight on labor in nulliparas. *Obstet Gynecol.* 1990;76:159–163.
11. Mor-Yosef S, Samueloff A, Modan B, Navot D, Schenker JG. Ranking the risk factors for cesarean: logistic regression analysis of a nationwide study. *Obstet Gynecol.* 1990;75:944–947.
12. Harnett Sheehan K. Caesarean section for dystocia: a comparison of practices in two countries. *Lancet.* 1987;i:548–551.
13. Hewson D, Bennett A, Holliday S, Booker E. Childbirth in Sydney teaching hospitals: a study of low-risk primiparous women. *Community Health Stud.* 1985;10:195–202.
14. Cary AJ. Intervention rates in spontaneous term labour in low risk nulliparous women. *Aust NZ J Obstet Gynaecol.* 1990;30:46–51.
15. Haynes de Regt R, Minkoff HL, Feldman J, Schwarz RH. Relation of private or clinic care to the cesarean birth rate. *N Engl J Med.* 1986;315:619–624.
16. Klein M, Lloyd I, Redman C, Bull M, Turnbull AC. A comparison of low-risk pregnant women booked for delivery in two systems of care: shared-care (consultant) and integrated general practice unit. I. Obstetrical procedures and neonatal outcome. *Br J Obstet Gynaecol.* 1983;90:118–122.
17. Yudkin PL, Redman CWG. Caesarean section dissected, 1978–1983. *Br J Obstet Gynaecol.* 1986;93:135–144.
18. Gordon D, Milberg J, Daling J, Hickok D. Advanced maternal age as a risk factor for cesarean delivery. *Obstet Gynecol.* 1991; 77:493–497.
19. Goodfellow CF, Hull MGR, Swaab DF, Dogterom J, Buijs RM. Oxytocin deficiency at delivery with epidural analgesia. *Br J Obstet Gynaecol.* 1983;90:214–219.
20. Thorp JA, Parisi VM, Boylan PC, Johnston DA. The effect of continuous epidural analgesia on cesarean section for dystocia in nulliparous women. *Am J Obstet Gynecol.* 1989;161:670–675.
21. Dickersin K. Pharmacological control of pain during labour. In: Chalmers I, Enkin M, Keirse MJNC, eds. *Effective Care in Pregnancy and Childbirth. Vol. 2: Childbirth.* Oxford, England: Oxford University Press; 1989;913–950.

Physicians' Recommendations for Mammography: Do Tailored Messages Make a Difference?

Celette Sugg Skinner, PhD, Victor J. Strecher, PhD, MPH, and Harm Hospers, MA

ABSTRACT

Objectives. **Message tailoring, based on individual needs and circumstances, is commonly used to enhance face-to-face patient counseling. Only recently has individual tailoring become feasible for printed messages. This study sought to determine whether printed tailored recommendations addressing women's specific screening and risk status and perceptions about breast cancer and mammography are more effective than standardized printed recommendations.**

Methods. **Computer-assisted telephone interviews were conducted with 435 women, aged 40 to 65 years, who had visited family practice groups within the previous 2 years. Subjects were randomly allocated to receive individually tailored or standardized mammography recommendation letters mailed from physicians to patients' homes. Follow-up interviews were conducted 8 months later.**

Results. **Tailored letter recipients were more likely to remember and to have read more of their letters than standardized version recipients. After controlling for baseline status, tailored letter receipt was associated with more favorable follow-up mammography status for women with incomes below $26 000 and for Black women.**

Conclusions. **Tailored messages are a more effective medium for physicians' mammography recommendations; tailoring may be especially important for women of low socioeconomic status. (*Am J Public Health.* 1994;84:43-49)**

Introduction

Breast cancer affects about 1 in 10 women and is one of the three most frequent causes of death among women in the United States.[1] Regular mammography screening has demonstrated effectiveness in reducing breast cancer mortality.[2-11] Although nearly all health organizations recommend mammography screening[12-19] and rates have been increasing,[20,21] most women are still not screened regularly.[21,22] Women of lower socioeconomic status are least likely to have regular mammograms.[20,23-27]

Most identified factors influencing women's mammography decisions can be classified as variables from the Health Belief Model,[28,29] which purports that perceptions about benefits and barriers associated with a health behavior, perceived personal risk and severity of the relevant health threat, and various cues to action work together to influence a person's likelihood of taking a preventive health action.

It is also likely that with regard to mammography as well as other health-related behaviors,[30-32] individuals move through a series of motivational and behavioral stages in which the behavior is not considered ("precontemplation"), the behavior is contemplated but not yet acted upon ("contemplation"), an initial behavior change is made ("action"), and the behavior is maintained over time ("maintenance") or the person relapses from action or maintenance to a previous stage. Research with other behaviors has found utility in considering health-related behavior changes in these categories[32-36] and has found a broad range of psychosocial determinants predicting movement from one stage to another.

Among factors correlated with mammography status are women's beliefs about breast cancer and the ability of mammography to detect early breast cancer[23,25,37-40] and perceived barriers to mammography screening, such as cost, concern about radiation exposure, or fear of finding cancer.[25-27,39-43] A physician's recommendation, however, has emerged as an even stronger predictor of mammography than women's perceptions about the procedure.[21,26,39,42-44]

Therefore, it is likely that mammography recommendations in the form of letters mailed from physicians to patients' homes would have an impact on women's propensity to be screened. We used computer technology to assemble individualized letters tailored according to women's responses in baseline interviews. Message adaptation, based on individual needs and circumstances, is commonly used to enhance the effectiveness of face-to-face patient counseling,[45,46] but only recently has such individual tailoring become feasible for printed messages. We sought to determine whether printed tailored recommendations addressing women's specific screening and risk status and their perceptions about breast cancer and mammog-

At the time of the study, Celette Sugg Skinner and Victor J. Strecher were with the Department of Health Behavior and Health Education, School of Public Health, University of North Carolina at Chapel Hill. Harm Hospers is with the Department of Health Education, University of Limburg, Maastricht, the Netherlands.

Requests for reprints should be sent to Celette Sugg Skinner, PhD, Division of Radiology Research, Washington University Medical School, 510 S Kingshighway, St Louis, MO 63110.

This paper was accepted June 3, 1993.

Editor's Note. See related editorials by Shapiro (p 10), and Zapka (p 12) in this issue.

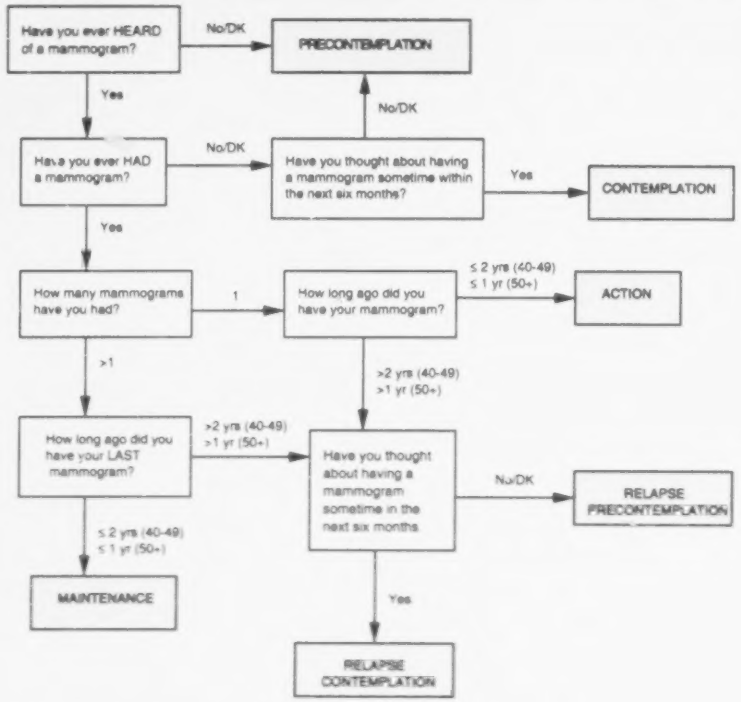


FIGURE 1—Mammography stage diagram.

raphy are more effective than standardized printed recommendations.

Methods

The eligible study population was a random sample of women aged 40 through 65 years who had visited one of two North Carolina family practice groups within the previous 2 years, had telephones, and had never been diagnosed with breast cancer. Study subjects were interviewed at baseline and randomly allocated to receive individually tailored or standardized letters. Whereas the general letter delivered a standardized message to all recipients, tailored letters addressed individual beliefs and barriers associated with mammography, as well as breast cancer risk factors and screening status. Follow-up interviews 8 months later reassessed the subjects' screening status and perceptions and measured letter recall and readership. Data were collected via computer-assisted telephone interviews.

The eligible study population consisted of 899 patients from both practices. Fifty-eight women (6%) refused to be interviewed, 10 interviews were terminated (1%), and 334 (37%) patients were never reached after an average of 7 attempts over 10 days. Two hundred patients from practice A and 297 from practice B were enrolled. Samples from the two practices were similar in racial and age distributions but differed somewhat in income and education levels. To minimize effects of differences, half of the subjects from each practice were randomly allocated to receive tailored versus standardized letters.

Baseline interviews asked the women about their mammography stage, their knowledge and beliefs about breast cancer and mammography, their objective and perceived risk status, and barriers to screening. Descriptions of measures follow.

Stage Determination (Likelihood of Taking Action)

We adapted a stage model of smoking cessation developed by Prochaska and DiClemente.[30-32] On the basis of their responses to several questions, women were classified as precontemplators, contemplators, actors, or maintainers. Figure 1 depicts the measures by which stages were defined.

The respondents were asked whether they had heard of mammograms and, if so, whether they had had one or more. All respondents who had heard of mammography but had not yet been screened were asked whether they had thought about being screened in the 6 months immediately following the interview. Respondents who had never heard of mammography or who had not thought about having a mammogram in the following 6 months were *precontemplators*. Those who had thought about having mammograms in the following 6 months were *contemplators*.

Following study physicians' recommendations and the American Cancer Society guidelines for frequency of mammography screening, women aged 50 years and older were considered due for repeat mammograms if their last mammograms had been more than a year ago and women aged 40 through 49 were considered due for rescreening if more than 2 years had passed since their last mammograms. Those who had not thought about having another mammogram in the following 6 months were *relapse precontemplators*, whereas those who had thought of being rescreened in the following 6 months were *relapse contemplators*. Those not yet due for rescreening were *actors* (had had one mammogram) or *maintainers* (had had two or more mammograms).

Beliefs about Benefits and Barriers Associated with Mammography

The interview assessed the subjects' beliefs about the ability of mammograms to detect early breast cancers and about the likelihood of curing breast cancers detected at an early stage. Responses were measured on 5-point scales.

Perceived barriers were elicited by the open-ended question "What would keep you from having a mammogram in the next 6 months?" Additionally, closed-ended questions assessed the extent to which respondents would be bothered by cost, discomfort, fear of finding cancer, and concern about radiation.

The Mammography Letters

A library of texts and graphics files was created to address varying beliefs, mammography stages, risk factors, and barriers. Messages were pretested in physicians' waiting rooms among patients not included in the sample. The brief message texts were written in a simple, popular style and could occur in any possible text combination. Desktop publishing techniques were used to enter message texts in predefined locations in the letter layout. Specifically, interview data from women in the experimental group were transferred from an SAS[47] data set to an ASCII file that delineated which text version each woman should receive. Data in each woman's ASCII file were then read by a Microsoft

Word[48] program that inserted message texts appropriate for her. A combination of messages was created to specifically address each woman's beliefs about mammography and breast cancer, mammography stage, risk status, and perceived barriers. More than 391 000 different tailored letters were possible. Every woman's letter was different from every other woman's unless the women had given identical responses in the baseline interview.

Five months after the baseline interview, the subjects received tailored or standardized letters according to the previous random assignment. Each letter was accompanied by a cover letter on a physician's letterhead with a digitized signature. Envelopes bore physicians' logos, address labels, and commemorative stamps.

The standardized message typifying usual care communications was an adaptation of the 1987 surgeon general's mammography letter.[49] To minimize differences in appearance, standardized and tailored letters were printed in the same typeface on the same stationery and were similar in length. Both had a picture of a woman, but tailored letters had digitized photographs tailored by race, accompanied by captions tailored by stage.

Whereas the standardized letters addressed a variety of factors that *might* be relevant to recipients, the tailored letters addressed only those that, according to the baseline interviews, *were* relevant to the recipients. For instance, tailored messages reported the risk factor of age only for women who were older than 50 and the risk factor of family history only for women who had a relative with breast cancer.

The tailored letters contained individual-specific information but were framed to sound as though they were written for general audiences. They were written as reports of findings from the telephone interviews in which all letter recipients had participated. For example, instead of "Since you, personally, have never even thought about having a mammogram . . . ," the messages read, "Many women surveyed have never even thought about having a mammogram. . . ."

Follow-Up Interviews

Second telephone interviews—conducted 3 months after the letters were mailed and 8 months after the baseline interviews—measured letter recall and readership and reassessed respondents' stage, perceived barriers, knowledge, and beliefs.

Recall of letter receipt was assessed in both unaided and aided format. The unaided question asked whether respondents had recently received any health information in the mail from the study physicians. If respondents did not remember or were unsure, they were prompted with a reminder about the letters and then asked again. For those who remembered the letters, readership was assessed on a 4-point scale ranging from "Read all of the information" to "Didn't read any of it."

Eight study subjects whose mammography recommendation letters had been returned to sender and for whom no new addresses could be found were dropped from the sample between baseline and follow-up. Of the remaining 489 subjects interviewed at baseline, 435 (89%) were reinterviewed at follow-up (170 women from practice A and 265 from practice B). Twenty-four subjects (5%) refused the second interview, 26 (5%) were never reached after an average of 7 attempts over 10 days, and 4 (1%) interviews were terminated.

Analyses

Assignment to experimental (tailored) or control (standardized) group was an independent variable for all analyses. Follow-up data were used to assess letter recall and readership. Both baseline and follow-up data were used to assess behavioral changes. Rather than conducting separate analyses for women who did and did not recall their letters or for women who did and did not read any of the information, in analyses for the full sample we controlled for letter recall and readership in assessing changes in mammography screening stage. Multivariate analyses were used to examine the main effects of the intervention on changes in mammography stage and on having a mammogram. Effects of the intervention were also examined within demographic groups (race, income, education, and age) for which mammography stage differed significantly at baseline. These demographic variables were entered into multivariate analysis models as interaction terms.

Results

The sample's demographic make-up was as follows: 53% were in their 40s, 84% were White (the remaining 16% were Black), and 68% were employed. Ten percent had not finished high school, 47% had at least attended college, and 16% had done graduate work. Forty percent had household incomes of $36 000 or more, 17% had incomes of $26 000 to $36 000, and 43% had incomes below $26 000. Bivariate analyses revealed no significant differences in any demographic characteristics, in family history of breast cancer, or in baseline mammography stage between experimental and control groups.

Recall and Readership

Most women (63%) remembered receiving their letters; 54% remembered without being prompted and an additional 9% remembered after a reminder. Only 17% of the women who remembered receiving a letter did not read any of the information. Whereas 36% of those who remembered receiving a letter read only some of the content, 23% read most and 24% read all of it. Among the 226 women who read at least some of the content, 30% said the information was very interesting and 59% found it at least somewhat interesting.

Bivariate analysis revealed that women who received tailored letters were more likely to remember them than were standardized letter recipients ($P < .05$). Further, among the 272 women who remembered the letters, tailored letter recipients were significantly more likely to read more of the content ($P < .01$). More than half (53%) of the tailored letter recipients who recalled the letter reported reading all or most of it, compared with 40% of standardized letter recipients.

Although women with higher educational levels were more likely to remember receiving the letters ($P < .01$), among those who read at least some of their letters, more well-educated women were less likely to report interest in the content ($P < .01$). Also, White women were more likely to remember receiving the letters ($P < .05$), but among women who remembered and read at least some of the letters, Black women reported more interest in the content ($P < .01$). A baseline perception of being at elevated risk for breast cancer was significantly related to reading more of the letter ($P < .01$). However, tests for interactions revealed no significant effects of these correlates and letter type on recall, readership, or interest.

Mammography Stage Movement

The study categorized respondents into mammography stages. Precontemplators, whether they had never had a mammogram or were due for rescreening, needed mammograms but had not thought about having them in the following 6

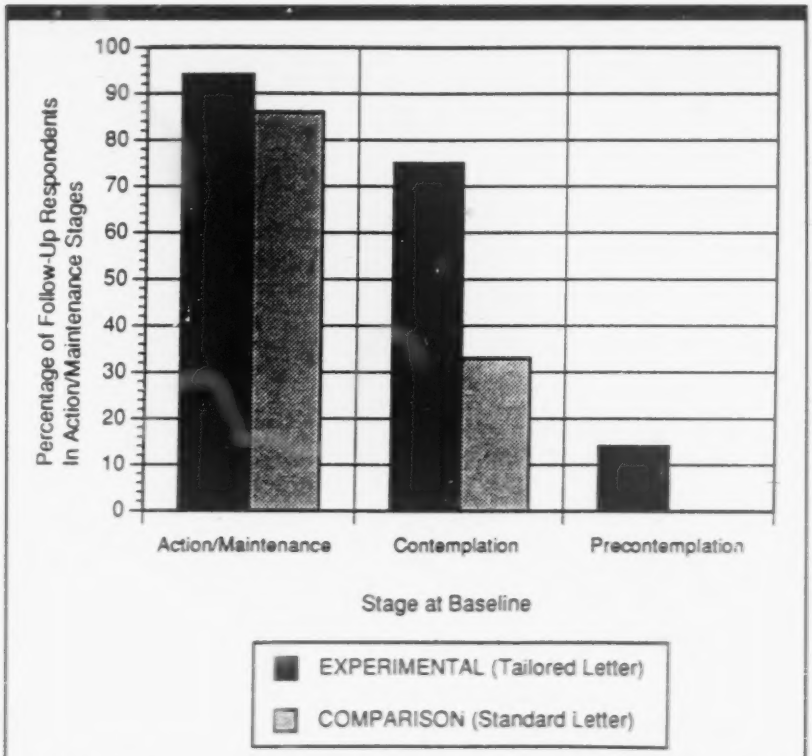


FIGURE 2—Movement toward action/maintenance stage of mammography use (Black women, n = 70).

months. Contemplators, also due for mammography, had thought about having mammograms at some time in the following 6 months. Actors and maintainers (grouped together for these analyses) were not due for repeat screening within the next 6 months. Within these stages, the possibility of forward and backward movement is apparent. The goal of the intervention was to move women into the action/maintenance category and keep them there by preventing relapse.

At baseline, 64% of all women in the sample had had mammograms within the recommended interval and were not due for rescreening. Women were more likely to have had recent mammograms at baseline if they were White ($P < .05$), had higher incomes ($P < .001$), and had higher education levels ($P < .01$). Thirty-five percent of White women, compared with 43% of Black women, were due for mammography screening at baseline. Only 28% of women with incomes over $26 000, compared with 47% of women with lower incomes, were due for screening at baseline. Forty-three percent of women who did not attend college, compared with 29% of those who did, were due for mammography at baseline. Age was associated with screening; women in their 40s were less likely than women aged 50 years and older to be due for a mammogram ($P < .001$).

The rate of women who had had recent mammograms increased from 64% at baseline to 68% at follow-up. Of the 152 women who were due at baseline, 37% had had mammograms by follow-up. Among those due for screening at baseline, more of the tailored letter recipients (44%) than the standardized letter recipients (31%) had had mammograms by follow-up. However, the difference was not statistically significant ($P = .16$).

A woman could progress from precontemplation (not thinking about having a mammogram in the next 6 months) to contemplation (thinking about having a mammogram in the next 6 months) without actually having a mammogram. Conversely, she could regress from contemplation to precontemplation. Actors and maintainers could relapse to precontemplation or contemplation stages if, by the follow-up interview, they had become overdue for rescreening. The majority of subjects (71%) did not change stage between baseline and follow-up (14% advanced one stage, 4% advanced two stages, 8% moved one stage back, and 3% regressed by two stages). There were no significant main effects for stage movement by letter type.

Analyses were conducted to determine whether the intervention had a significant effect on mammography within demographic subgroups. Letter type was not differentially associated with stage movement for different education or age groups. For Black and low-income women, however, receipt of tailored letters, compared with standardized letters, influenced mammography stage improvement and mammography screening rates. After controlling for stage at baseline, significant race × intervention ($P < .05$) and income × intervention ($P < .01$) interactions were found.

To determine the nature of the interaction effects, we ran four separate models for lower-income, higher-income, Black, and White women. Letter type was significantly associated with stage movement among Black women ($P < .05$) but not among White women ($P = .47$). Twenty-seven percent of Black women who received tailored letters, compared with 8% in the standardized letter group, exhibited forward stage movement at follow-up. Further, letter type was significantly associated with stage movement among women with incomes below $26 000 ($P < .01$) but not among higher-income women ($P = .30$). In the lower-income group, more tailored letter recipients moved forward (28% vs 17% of standardized letter recipients) and fewer moved backward (8% vs 15% of standardized letter recipients).

Figure 2 depicts percentages of Black women in the action/maintenance stage at follow-up, delineated by baseline stage and letter type. Figure 3 depicts corresponding breakdowns for women with household incomes below $26 000. For both Black and low-income women, baseline precontemplators and contemplators were more likely to have had mammograms at follow-up if they had received tailored rather than standardized letters. Similarly, among both groups, baseline action/maintainers who received tailored letters were less likely to have relapsed than were standardized letter recipients.

Discussion

In this study we sought to determine whether printed physicians' recommendations for mammography would be more effective if they were tailored to individual women's specific perceptions about mammography and breast cancer, their breast cancer risk factors, and their mammo-

graphy screening status. The results indicate that tailored letters were a more effective medium for delivering the message: tailored letters were more likely to be remembered and, among women who remembered the letters, were more thoroughly read.

There are several explanations for why the tailored letters were better remembered and more thoroughly read. Although the tailored and standardized letters were designed to be equally attractive, tailored letters may have had a greater chance of capturing attention at first glance because of their tailored pictures, captions, and headlines. Tailoring allowed for a specificity in the letters' eye-catching elements that was not possible in the standardized version. For example, the caption for relapse precontemplators' letters read, "I've had a mammogram but I never even thought about having another one."

Had the standardized letter contained such specificity, some of its elements would have been irrelevant, and possibly off-putting, for a portion of its recipients. Further, the specificity of content allowed in the tailored letter kept recipients from reading information that did not apply to them. In other words, tailoring enhanced relevance. For instance, because only risk factors relevant to the recipient were printed, a woman could not read a list of three risk factors and feel consoled by the fact that only one applied to her. Tailored letters may have simply received enhanced attention, which in turn led to forward stage movement. In the context of the Health Belief Model,[29] they may have been more effective cues to action.

Most women (64%) entered the study with an up-to-date mammography status. This rate increased to 68% by follow-up. Although more women in the tailored than the standardized group had had recent mammograms at follow-up, the difference did not achieve statistical significance. Among Black and low-income women, whose recent screening levels were lower at baseline, tailored letter recipients were more likely than standardized letter recipients to move toward mammography. In addition, those who received the tailored letters were less likely than standardized letter recipients to move backward or relapse.

This study has several limitations. For instance, its generalizability is limited by the exclusion of women without telephones. In addition, there is a possibility that women who were never reached were different from those who were contacted.

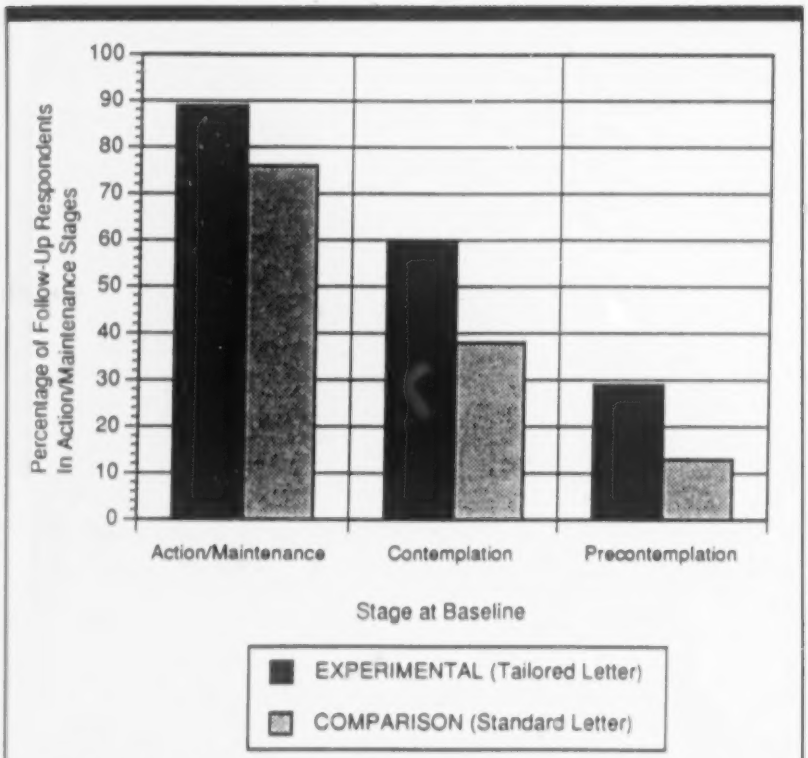


FIGURE 3—Movement toward action/maintenance stage of mammography use (women with household income < $26 000; n = 171).

Women not reached may have been out of town during the calling period, may have been generally busier and at home less often than those who were reached, or may have screened their calls via answering machines and chosen not to answer. It seems unlikely, however, that any differences between those who were and those who were not reached would be associated with the effect of receiving a tailored or standardized letter.

For the sample as a whole, higher-than-expected mammography rates at baseline resulted in limited statistical power to detect postintervention differences between tailored and standardized groups. Among subgroups, it is unclear whether there were aspects of the tailored letters that were particularly effective for Black and low-income women or whether there was a ceiling effect for White and higher-income women because their percentage of possible change was limited by elevated baseline rates. Either way, the most important finding of this study is its demonstration that tailored physicians' recommendations effected health behavior change among hard-to-reach populations. Minority and low-income women have historically had lower mammography rates than White women and those with higher incomes, and they remain underscreened although screening rates continue to rise in the general population. In fact, the mammography gap between higher-income White women and lower-income minorities may be widening.[20] Therefore, an intervention found to be effective among lower-income and minority groups is especially timely and important.

It is somewhat surprising that the effect of the tailored letters was significant for Black and low-income women, because print media are not generally expected to be as effective among lower socioeconomic groups as among higher socioeconomic groups. To combat this disadvantage, letters were targeted toward lower reading levels by means of simple vocabulary and clear, one-sided arguments. In effect, the intervention delivered a low-socioeconomic-status message through a high-socioeconomic-status channel. This may explain why Black and low-income women were among the least likely to have read their letters but, for women who read at least part of the letters, the most likely to report interest in the content.

The effect of tailored recommendations cannot, in this study, be disentangled from the personal nature of letters coming

from physicians to women's homes. It is difficult to determine whether the advantage in tailoring would also be present for messages delivered via other media. Women of low socioeconomic status may have been persuaded more by the fact that their physicians sent them a letter than they would have been by an even easier-to-understand presentation in the physician's office.

The next phase in tailoring research should compare different types and amounts of tailoring and different media channels. For instance, interactive computer programs employing graphics, sound, and touch screens could deliver messages to low-literacy women who are not able to read letters. Such messages could allow for additional tailoring. Not only could more content be tailored, but presentation factors such as type size, graphics, reading level, and sophistication of arguments or message framing could be manipulated.

Future research should also explore whether effects of tailored physicians' recommendations are similar for patients in public health clinics, where physician–patient relationships may be very different from those in family practice settings. Even if tailored recommendations are found to be more effective in moving women toward mammography, they may not provide women with the skills needed to negotiate the medical system and actually obtain mammograms. □

Acknowledgments

This research was supported by Public Health Service Grant 1 R29 HL41886 from the National Heart Lung and Blood Institute to Dr Strecher and a Cancer Control Research Grant from the Lineberger Comprehensive Cancer Center to Dr Skinner.

The authors thank Dirk-Jan Den Boer, PhD; Sara Banks, MPH; Margaret Hasty, MPH; and Michael Tyler, MD, for their valuable contributions to this work.

References

1. American Cancer Society. Cancer statistics 1991. *CA Cancer J Clin.* 1991;41(1):19–36.
2. Shapiro S. Evidence on screening for breast cancer from a randomized trial. *Cancer.* 1977;39:2772–2782.
3. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *J Nat Cancer Inst.* 1982;69:349–355.
4. Tabar L, Fagerberg CJG, Day NE, Holmberg L. What is the optimum interval between mammographic screening examinations? An analysis based on the test results of the Swedish two-county breast cancer screening trial. *Br J Cancer.* 1987;55:547–551.
5. Tabar L, Fagerberg CJG, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography. *Lancet.* 1985;1:829–832.
6. Collette HJA, Day NE, Rombach JJ, de Waard F. Evaluation of screening for breast cancer in a non-randomized study (the DOM Project) by means of a case-control study. *Lancet.* 1984;1:1224–1226.
7. Palli D, Del Turco MR, Buiatti E, et al. A case–control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). *Int J Cancer.* 1986;38:501–504.
8. Verbeek ALM, Hendriks JH, Holland R, Mravunac M, Sturmans F, Day NE. Reduction of breast cancer mortality through mass screening with modern mammography: first results of the Nijmegen project, 1975–81. *Lancet.* 1984;i:1222–1224.
9. Peeters PHM, Verbeek ALM, Hendriks JH, Holland R, Mravunac M. The predictive value of positive test results in screening for breast cancer by mammography in the Nijmegen programme. *Br J Cancer.* 1987;56:667–671.
10. Seidman H, Gelb S, Silverberg E, LaVerda N, Lubera JA. Survival experience in the Breast Cancer Detection Demonstration Project. *CA Cancer J Clin.* 1987;37:258–290.
11. Morrison AS, Brisson J, Khalid N. Breast cancer incidence and mortality in the Breast Cancer Detection Demonstration Project. *J Nat Cancer Inst.* 1988;19:1540–1547.
12. *Mammography Statement.* Washington, DC: American College of Obstetricians and Gynecologists; 1983.
13. American College of Physicians. Preventive care guidelines: 1991. *Ann Intern Med.* 1991;114:758–783.
14. American College of Radiology. Guidelines for mammography. *Am Coll Radiol Bull.* 1982;38:6–7.
15. American Medical Association Council on Scientific Affairs. Early detection of breast cancer. *JAMA.* 1984;252:3008–3011.
16. American Cancer Society. Mammography guidelines, 1983: background statement and update of cancer-related check-up for breast cancer detection in asymptomatic women age 40–49. *CA Cancer J Clin.* 1983; 33:255.
17. Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J.* 1986;134: 721–729.
18. *Working Guidelines for Early Cancer Detection.* Bethesda, Md: National Cancer Institute; 1987.
19. United States Preventive Services Task Force. Recommendations for breast cancer screening. *JAMA.* 1987;257:2195.
20. Zapka J, Stoddard AM, Constanza ME, Greene HL. Breast cancer screening by mammography: utilization and associated factors. *Am J Public Health.* 1989;79: 1499–1502.
21. National Cancer Institute Breast Cancer Screening Consortium. Screening mammography: a missed clinical opportunity? Results of the NCI Breast Cancer Screening Consortium and National Health Interview Survey studies. *JAMA.* 1990;264:54–58.
22. *Breast Cancer: A Measure of Progress in Public Understanding.* Washington, DC: US Dept of Health and Human Services, National Institutes of Health; October 1980. NIH publication 81-2291. Technical Report.
23. MacLean U, Sinfield D, Klein S, Harnden B. Women who decline breast screening. *J Epidemiol Community Health.* 1984;38: 278–283.
24. Coll PR, O'Connor PJ, Crabtree BF, Besdine RW. Effects of age, education and physician advice on utilization of screening mammography. *J Am Geriatr Soc.* 1989; 37:957–962.
25. Rutledge DN, Hartmann WH, Kinman PO, Winfield AC. Exploration of factors affecting mammography behaviors. *Prev Med.* 1988;17:412–422.
26. Tippy P, Falvo D, Woehlke P. Fee structure as a determinant of patient's choice to undergo mammography. *Fam Prac Res J.* 1989;9:43–51.
27. Howard J. Using mammography for cancer control: an unrealized potential. *CA Cancer J Clin.* 1987;37:2–19.
28. Hochbaum GM. *Public Participation in Medical Screening Programs: A Sociopsychological Study.* Washington, DC: US Dept of Health, Education, and Welfare; 1958. PHS publication 572.
29. Becker ME. The Health Belief Model and personal health behavior. *Health Educ Monogr.* 1974;2(4).
30. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol.* 1983;51:390–395.
31. Prochaska JO, DiClemente CC. Measuring process of change: applications to the cessation of smoking. *J Consult Clin Psychol.* 1988;56:520–528.
32. Prochaska JO, Goldstein MG. Process of smoking cessation: implications for clinicians. *Clin Chest Med.* 1991;12:727–734.
33. DiClemente CC, Hughes S. Stages of change profiles in outpatient alcoholism treatment. *J Subst Abuse.* 1990;2:217–235.
34. O'Connell D, Velicer WF. A decisional balance measure for weight loss. *Int J Addict.* 1988;23:729–750.
35. Rossi JS. Exploring behavioral approaches to UV risk reduction. In: Moshell A, Blankenbaker LU, eds. *Sunlight, Ultraviolet Radiation, and the Skin.* Bethesda, Md: National Institutes of Health; 1989:91–93.
36. Rakowski W, Jube CE, Marcus BH, Prochaska JO, Velicer WF, Abrams DB. Assessing elements of women's decisions about mammography. *Health Psychol.* 1991;11:111–118.
37. Calnan MW, Moss S, Chamberlain J. Explaining attendance at a breast-screening clinic. *Patient Educ Counseling.* 1985;7: 87–96.
38. Burack RC, Liang J. The acceptance and completion of mammography by older Black women. *Am J Public Health.* 1989; 79:721–726.
39. Lane DS, Fine HL. Compliance with mammography referrals: implications for breast cancer screening. *N Y State J Med.* 1983; 83(February):173–176.
40. Champion VL. The relationship of selected variables to breast cancer detection behav-

iors in women 35 and older. *Oncol Nurs Forum.* 1991;18:733–739.
41. Kruse J, Phillips D. Factors influencing women's decisions to undergo mammography. *Obstet Gynecol.* 1987;70:744–748.
42. Lerman C, Rimer B, Trock B, Balshem A, Engstrom P. Factors associated with repeat adherence to breast cancer screening. *Prev Med.* 1990;19:279–290.
43. Wolosin RJ. The experience of screening mammography. *J Fam Pract.* 1989;29:499–502.
44. Smith RA, Haynes S. Barriers to screening for breast cancer. *Cancer.* 1992;69:1968–1978.
45. Bartlett EE. The contributions of behavioral science to patient education practice: a review. In: Ruffini JL, ed. *Advances in Medical Social Science.* Vol. 2. New York, NY: Gordon and Breach Science Publishers; 1984.
46. DeVellis B, Blalock S, Hochbaum G, Hahn P. Evaluation of a problem-solving intervention for patients with arthritis. *Patient Educ Counseling.* 1988;11:29–42.
47. *SAS—Statistical Guide for Personal Computers, Version 6.* Cary, NC: SAS Institute Inc; 1987.
48. Microsoft Word Document Processing Program, Version 4.0 for the Apple Macintosh. Redmond, WA: Microsoft Corporation; 1989.
49. Koop CE. *A Message from the Surgeon General.* Washington, DC: US Dept of Health and Human Services; 1987.

The Collaborative Registry of Smoking Cessation Trials Issues a Request for Help

Over the past 10 years there has been a significant increase in the number of randomized controlled trials that compare the effectiveness of different interventions in smoking cessation. Recently, a number of researchers have been approached to obtain information about current trials for a prospective registry of planned or ongoing trials in the area of smoking cessation. Although the initial response has been encouraging, it is possible that some people have been inadvertently missed. To ensure that a registration form is completed by the principal investigator for all relevant trials, we are seeking the help of all researchers.

To be eligible for inclusion in the registry, a trial must be unpublished and include at least two groups; allocation to the groups must be by either a random or quasi-random method (e.g., alternation, year of birth, etc.); and the trial must be related to an aspect of smoking cessation. Trials examining abstinence rates, relapse prevention, withdrawal symptoms, training or encouraging health professionals in smoking cessation techniques, or any aspect of smoking cessation research are all eligible for inclusion. If you are in doubt as to whether a trial is suitable for inclusion, we suggest you still complete a registration form.

Once the register has been assembled, a copy will be distributed to all contributors, as well as published in summary form on an annual or biannual basis. The registry will not collect any trial result data or participant information, although the existence of such a register may facilitate the establishment of collaborative groups who wish to undertake more detailed systematic reviews, similar to those undertaken in other fields.

Trial registration forms are available on request from The Collaborative Registry of Smoking Cessation Trials, General Practice Research Group, Gibson Building, Radcliffe Infirmary, Oxford OX2 6HE, United Kingdom; tel +44-865-319 124; fax +44-865-310 545.

The coordinators would also appreciate being informed of any completed but unpublished smoking cessation trials. No special form is provided for this purpose; however, any information that researchers can provide will assist in updating our current register of completed trials and ensure its comprehensiveness.

Mammography Screening: How Important Is Cost as a Barrier to Use?

Nicole Urban, ScD, Garnet L. Anderson, PhD, and Susan Peacock, BA

ABSTRACT

Objectives. Recent legislation will improve insurance coverage for screening mammography and effectively lower its cost to many women. Although cost has been cited as a barrier to use, evidence of the magnitude of its effect on use is limited.

Methods. Mammography use in the past 2 years among women aged 50 to 75 residing in four suburban or rural counties in Washington State was estimated from 1989 survey data. Logistic regression analysis was used to estimate the odds ratio of mammography use as a function of economic and other variables. Within a residential area, averages were used to measure the market price of mammography and the time cost to obtain a mammogram.

Results. Use was lower among women who faced a higher net price or who preferred to obtain a mammogram during weekend or evening hours and higher among women with higher incomes. Visiting no doctor regularly and smoking were predictors of failure to use mammography.

Conclusion. The effects of economic variables on mammography use are important and stable across subsets of the population, but they are modest in size. (*Am J Public Health.* 1994;84:50–55)

Introduction

The burden of breast cancer in the United States is great and is not likely to decrease in the near future.[1] Although one out of nine US women can expect to be diagnosed with breast cancer during her lifetime, and regular screening by mammography among women aged 50 or older has been shown to reduce breast cancer mortality by 30% or more,[2,3] surveys done between 1988 and 1989 suggest that fewer than 42% of US women aged 50–75 have had mammography in the past year and that over 25% of women aged 50–75 have never had a mammogram.[4]

The reasons most frequently cited by women for never having had mammography are that they do not have any problems with their breasts, and that their doctors never recommended it.[4–10] A recent study of use of mammography by primary-care physicians found cost and lack of insurance coverage to be major deterrents to use.[10] In studies of women's use of mammography in which cost has been measured, cost has been identified as a factor influencing use,[5,8] and the role of insurance has been found to be important.[5,9]

If cost is a significant barrier, mammography use can be expected to increase in the next decade because recent legislation in many states mandates that screening mammography be included among covered services in private health insurance policies,[11] and many of the remaining states have passed or are considering legislation mandating coverage of screening mammography by insurers. Moreover, effective January 1, 1991, biennial screening mammography has been reimbursable by Medicare. Until then, Medicare reimbursed 80% of reasonable and customary charges for diagnostic mammography once the deductible was met, but benefits for screening mammography were not provided.

This paper examines the role of economic variables, including insurance coverage, in the use of mammography among women aged 50–75 in four counties in Washington State before implementation of legislation that improved insurance coverage. Data are presented describing the perceptions of women in 1989 regarding their insurance coverage for screening mammograms and the influence of cost and insurance coverage on use. The bivariate relationships among recent mammography use and insurance coverage, mammography price, and income are reported for women aged 50 to 64 and for women aged 65 to 75.

An economic behavioral model, the theory of demand, suggests that the price of mammography is a key variable. Total price to the consumer is the net money price plus the time cost. The net money price is equal to the price charged for the mammogram less the amount paid by insurance, or the price charged multiplied by the coinsurance rate. The time cost is the total time required to obtain a mammogram. Other things being equal, as the net price of a screening mammogram falls, regular screening mammography will be used by more women. The theory of demand suggests as well that (1) use will rise with income; (2) use will rise with a fall in the time cost; (3) as the net money price falls, sensitivity to time cost will rise; and (4) price sensitivity will be greater among

The authors are with the Cancer Prevention Research Program, Fred Hutchinson Cancer Research Center, Seattle, Wash.

Requests for reprints should be sent to Nicole Urban, ScD, Fred Hutchinson Cancer Research Center, 1124 Columbia St MP-702, Seattle, WA 98104.

This paper was accepted April 29, 1993.

low-income women for whom price represents a larger proportion of disposable income. The results of multivariate analyses to test these hypotheses are reported.

Methods

Data and Variable Definitions

Data on mammography use were obtained from a telephone survey conducted in spring 1989 of women aged 50–75 residing in four communities. The communities were selected on the basis of their comparability with respect to size, sociodemographics, and proximity to an urban area from among all the suburban and rural counties in Washington State with at least 10 000 women aged 50 and above. To be eligible for the survey, women had to be between the ages of 50 and 75, had to have lived in their county of residence for 2 or more years, and had to have no history of breast cancer. Approximately 97% of the target population had telephones; of those without telephones, many would have been ineligible for the survey due to lack of 2-year residence in the community. A total of 1538 women were interviewed through a random-digit dial survey technique that samples numbers proportional to size (for residential lines) within a dialing prefix, producing a self-weighted stratified sample for the geographical areas.[12] The response rate to the survey was 72%, calculated as the number of completed surveys divided by the estimated number of eligible women sampled, which was quite good relative to the response rates reported in similar surveys.[4] The denominator included an estimate of the number of eligible women in households at telephone numbers that were sampled but were not reached (n = 192), as well as all eligible women reached (n = 1937). To avoid selection bias, the introduction to the survey described it as relating to the health of women rather than to mammography specifically. The questions used to elicit relevant information from women respondents are available from the authors on request.

The use of mammography was defined as one or more mammographic studies in the past 2 years, based on the woman's self-report. Validation was not undertaken because validity of self-reported mammography use has been found to be very good in other studies.[13] Although interest was primarily in screening mammography, no attempt was made to distinguish between screening and diagnostic mammograms because the distinction between them is insufficiently clear among physicians to be recognized reliably by women.[14] Of women who had ever had a mammogram, 82.4% reported that their last mammogram was for a routine check-up rather than for a current or previous breast problem. We are therefore confident that the majority of mammograms were for screening purposes.

Specification and Estimation of the Multivariate Models

Variables were selected for inclusion in the analytic model on the basis of microeconomic theory and the health economics literature. A single-equation model was specified that related use of mammography in the past 2 years to (1) economic variables and (2) other determinants of demand that reflect the value that women place on mammography as a contributor to their health.[15] Economic variables included in the model were the out-of-pocket cost to the woman of a screening mammogram (the net money price), the cost in time that is incurred to obtain the mammogram, whether or not weekdays are inconvenient for obtaining the mammogram, and household income.

The net money price was defined as the total charge for the mammography multiplied by the proportion of that charge paid by the woman rather than by her insurance company. For women with insurance that pays all of the cost of mammography, the net price is zero. The time cost was defined as the total time a woman would need to obtain the mammogram, including travel time to the facility, waiting time, and the time required to receive services. Measurements of the net money price and the time cost are described below. Inconvenience was measured as a woman's preference for evenings or weekends over weekdays to obtain the mammogram. Income was measured as reported total household income.

Theory suggests that demand is a function of the value of the good or service to the consumer, as well as a function of prices and income. The value of breast cancer screening to a woman was assumed to depend on her risk of breast cancer, measured by her age and family history of breast cancer; her valuation of her breast health, measured by her monthly practice of breast self-examination; and her valuation of her health more generally, which for lack of a better indicator was measured by her current smoking habits. Screening mammography can be viewed as an investment in health. Because microeconomic theory suggests that consumers with more education can be expected to invest more in their health,[15] years of education completed were also included in the model.

Because a woman may depend on her physician to assess the value of screening mammography for her,[16] a set of dummy variables indicating the type(s) of physician(s) she visits regularly were constructed. Based on bivariate analyses, these variables were reduced to two: an indicator that the woman regularly visits a gynecologist and an indicator that she does not visit any provider regularly. Because the probability of referral for mammography by a physician may increase with exposure to physicians, the average number of annual visits to providers whom the woman sees regularly was included as well.

Logistic regression analysis was used to estimate the model from the survey data obtained from the women. Of the 1538 women surveyed, there were 1528 usable interviews, of which 1377 had sufficiently complete data to be included in the demand analyses. The dependent variable was set equal to 1 if the woman reported having a mammogram within the last 2 years and set equal to 0 otherwise. The statistical software package EGRET[17] was used to estimate the model. Dummy variables were included for three of the four counties to account for county effects and any intraclass correlation among women in the same county.

An equation was also estimated in which the economic variables were omitted, in order to test their importance as a group; and two additional equations were estimated to test the interaction effects of economic variables with the economic conditions of the women. First, interaction effects were estimated for women with insurance covering more than half of the cost of screening mammography, which permitted a test of the hypothesis that sensitivity to the time cost is greater among well-insured women. Second, interaction effects were estimated for women with household incomes above $15 000, which yielded a test of the specific hypothesis that sensitivity to the money price is greater among low-income women.

Estimation Issues in the Multivariate Analyses

Missing values for economic variables were an important source of potential selection bias in these analyses. Among the 1528 usable interviews, the prevalence of missing values for reported

TABLE 1—Percentage of Women Aged 50–75 y Who Reported That Their Health Insurance Paid for Screening Mammography, by Age

Does Your Health Insurance Pay for Routine Screening Mammograms?	Age, %		
	50–64 y (n = 911)	65–75 y (n = 609)	All Ages (n = 1520)
Yes, all	31.7	31.2	31.5
Yes, partially	23.5	19.1	21.7
No, or no health insurance	28.1	21.4	26.0
Don't know	15.7	28.4	20.8

TABLE 2—Percentage of Women Aged 50–75 y Who Had a Mammogram within 1 or 2 Years of Interview, by Insurance Coverage and Age

Time Since Last Mammogram	Insurance Coverage for Screening Mammography							
	Covers All		Covers Part		Covers None[a]		Insurance Unknown	
	Age 50–64 y (n = 84)	Age 65–75 y (n = 186)	Age 50–64 y (n = 213)	Age 65–75 y (n = 114)	Age 50–64 y (n = 264)	Age 65–75 y (n = 130)	Age 50–64 y (n = 143)	Age 65–75 y (n = 173)
Within 1 year	52.1	52.2	48.8	46.5	39.0	45.4	18.2	11.0
Between 1 and 2 years	21.8	16.1	24.4	18.4	19.3	21.5	7.7	15.0
Over 2 years	11.3	13.4	14.1	15.8	16.3	10.0	20.3	19.7
Never	14.8	18.3	12.7	19.3	25.4	23.1	53.9	54.3

[a]Includes women with no insurance coverage and women with health insurance that does not cover screening mammography.

price, reported insurance coverage, reported time cost, and income were 19.4%, 26.5%, 9.7%, and 16.1%, respectively. Moreover, missingness for three of these economic variables was related to mammography use. Simple chi-square tests of independence between the variable indicating whether a woman had a mammogram in the last 2 years and the variables indicating whether or not she reported money price, insurance coverage, and time cost were highly significant ($P < .001$). Missingness for income was not related to mammography use.

Another potential source of bias was a possible correlation of reported price (in money, time, or both) and insurance coverage with the error term of the demand equation. Such a correlation would be introduced if women's recent use of mammography made them better informed regarding price and if poorly informed women systematically believed the price to be higher or lower than it actually was.

Predicted, rather than actual, prices and insurance coverage were therefore used to compute net price. Sixty-eight telephone number prefixes were used as indicators of approximate market areas. For each prefix area, the mammography price was estimated as the average price reported by women with that prefix who knew, or were willing to guess, the cost of screening mammography in their community. When there were fewer than three such women in a prefix area (a situation that occurred for 16 women), the overall average for the community was used. A dummy variable was included in the prediction equation indicating whether or not the woman had guessed, to allow for the possibility that these women systematically guessed higher or lower than women who claimed to know the price.

The procedure for predicting time cost was identical to that used for predicting money price, and an analogous approach was used to predict values for insurance coverage. All women were asked what kinds of health insurance they had. Fifty-six types of health insurance coverage were identified, and regression analysis (with dummy variables for each type of insurance) was used to estimate the proportion of the price paid by each insurer based on the survey responses of those women who answered the questions. These estimates were then used in place of reported insurance coverage in computing the net price as the product of the charge for mammography and the proportion paid by the woman. This procedure for imputing prices is an instrumental-variables approach that yields consistent estimates of the coefficients of interest.[18] It simultaneously computes missing values, precludes correlation of the regressor with the error term, and specifies demand as a function of the *market* price faced by the woman.

To avoid the exclusion from the sample of the women whose household income was not reported, a regression model based on geographic area as indicated by telephone prefix and sociodemographic characteristics of the woman (age, education, working status, and marital status) was estimated. Predicted incomes based on this equation were used to impute missing values.

Results

Of the 1528 usable interviews, 8 were excluded from all analyses because the age was missing. Approximately 40% of the remaining 1520 women were over the age of 65. The percentages of women, by age, who reported having health insurance that covered screening mammography are shown in Table 1. Over half of the women reported that they had insurance that paid all (31.5%) or part (21.7%) of the cost of a screening mammogram. The older women were slightly less likely than the younger women to report insurance coverage that paid all or part of the cost (50.3% vs 55.2%). The younger women were more likely than the older women to report no health insurance for screening mammograms (28.1% vs 21.4%), and the older women were more likely than the younger women not to know whether their insurance would pay for screening mammograms (28.4% vs 15.7%). Differences were statistically significant (χ^2 with 3 degrees of freedom [df] = 40.1, $P < .001$).

Bivariate relationships between mammography use and economic variables, based on the perceptions of the women themselves, are reported in Tables 2 and 3. Of the 1515 women reporting their mammography use, 40.5% had had a mammogram within the past year. Another 18.7% had had one between 1 and 2 years ago, and only 26.1% had never had a mammogram. The percentages of women who had a mammogram within 1 year, 2 years, more than 2 years ago, and never, by age and insurance coverage, are reported in Table 2. Over half of the women reporting that they did not know whether their insurance covered screen-

ing mammography had never had a mammogram: 53.9% of those under the age of 65 and 54.3% of those aged 65 or older, respectively. Of the women with full coverage, just over half had had mammography in the past year. About a quarter of the women who reported that their insurance did not cover any portion of the cost of screening mammography (or who reported no insurance at all) had never had a mammogram.

The percentages of women who had a mammogram within 1 year, 2 years, more than 2 years ago, and never, by age and income, are given in Table 3. Low-income women were much more likely never to have had a mammogram. Recent mammography (within 2 years) was reported most often among the younger women with incomes above $15 000 annually (68.7%) and least often among the younger women with incomes below $15 000 (42.5%). Age differences were not statistically significant in these analyses, but income and insurance were both significantly related to use when age was controlled (χ^2 with 1 df = 12.8 and 160.7, respectively; both $P < .001$).

Definitions for all variables used in the multivariate analyses are given in Table 4. A total of 1377 women had sufficiently complete data to be included. Means and standard deviations (or percentages for dichotomous variables) are shown for women who did (59.4%) and did not (40.6%) report having had a mammogram in the past 2 years. The statistical significance of the differences between the two groups is also shown for all variables. The average net price was predicted as $31 among users of mammography and $35 among nonusers of mammography. (The average price of mammography in the market areas where the women resided was reported to be $77, among the 1148 women who knew or were willing to guess the price. On average, among the 1032 women who knew whether their insurance covered screening mammography, 61% of the price was believed to be paid by insurance.) Users differed significantly from nonusers with respect to all variables except performance of breast self-examination and time cost. They differed from nonusers most notably with respect to whether they visited a gynecologist regularly (19.9% vs 6.1%), whether they reported that they visited no doctor regularly (3.4% vs 19.1%), and whether they were current smokers (15.0% vs 26.7%). Nonusers also reported lower incomes than users ($19 300 vs $26 300 annually), and a higher proportion of them reported

TABLE 3—Percentage of Women Aged 50–75 y Who Had a Mammogram within 1 or 2 Years of Interview, by Income and Age

Time Since Last Mammogram	Income <$15 000		Income ≥$15 000		Income Unknown	
	Age 50–64 y (n = 174)	Age ≥65 y (n = 229)	Age 50–64 y (n = 613)	Age ≥65 y (n = 252)	Age 50–64 y (n = 117)	Age ≥65 y (n = 122)
Within 1 year	27.0	34.5	47.5	42.1	36.8	35.3
Between 1 and 2 years	15.5	14.9	21.2	17.5	16.2	22.1
Over 2 years	16.7	13.1	13.7	15.5	18.0	17.2
Never	40.8	37.6	17.6	25.0	29.1	25.4

TABLE 4—Variable Definitions and Descriptive Statistics for Women Aged 50–75 y

	Users of Mammography (n = 818)		Nonusers of Mammography (n = 559)		
	Mean or %	SD	Mean or %	SD	P
Age ≥65: 1 if age was >65	36.8%		42.4%		.037
Married: 1 if married or living as married	74.1%		63.5%		<.001
Working: 1 if working full-time or part-time	33.3%		32.0%		.248
Education: mean years of education	13.2	2.2	12.7	2.3	<.001
Smoker: 1 if current smoker	15.0%		26.7%		<.001
Breast self-examination: 1 if performed monthly	50.9%		51.5%		.809
History: mean number of first-degree relatives with breast cancer	.174	.410	.095	.293	<.001
Physician visits: average number of annual visits made to physicians seen regularly	1.90	1.30	1.42	1.30	<.001
No doctor: 1 if no physician was visited regularly	3.4%		19.1%		<.001
Gynecologist: 1 if regularly visited gynecologist	19.9%		6.1%	.239	<.001
Net price: mean out-of-pocket cost of mammography, in $10 increments	$3.07	$1.70	$3.50	$1.71	<.001
Time cost: total hours set aside to obtain a mammogram	1.32	0.39	1.35	0.47	.181
Inconvenience: 1 if weekdays are less convenient than evenings and weekends to obtain mammography	9.0%		15.2%		<.001
Income: mean household income in $10 000 increments	$2.63	$2.08	$1.93	$1.79	<.001
No insurance: 1 if no health insurance	3.5%		10.4%		<.001
Medicare only: 1 if Medicare only	1.6%		7.3%		<.001
Health maintenance organization: 1 if insured through HMO	13.7%		9.3%		.011

that obtaining mammography on weekdays was inconvenient (15.2% vs 9.0%) and that Medicare was their only form of health insurance (7.3% vs 1.6%).

Results of the multivariate analysis are given in Table 5. As predicted by the theory of demand, use was lower among women who faced a higher net money price (odds ratio [OR] = 0.91 per $10 increase in net price, $P = .006$) or who preferred to obtain a mammogram during weekend or evening hours (OR = 0.53, $P = .001$), and higher among women with higher incomes (OR = 1.14, with a 14% increase associated with each $10 000 increment in total household income; $P < .001$). The effect of time cost was negative, as expected, but not statistically significant (OR = 0.86, with a decrease of 14% for each additional hour required; $P = .29$). As a group, the economic variables were highly statistically significant (likelihood ratio with 4 df = 34.0, $P < .001$).

The value of mammography to the woman was also important in determining demand. Use was higher among women with a family history of breast cancer (OR = 2.22, with more than a twofold increase for each first-degree relative with breast cancer; $P < .001$) and among women who visited a gynecologist regularly (OR = 2.59, $P < .001$). Women who visited no doctor regularly and women who were current smokers were less likely to have had mammography (OR = 0.24, $P < .001$ for no doctor; OR = 0.59, $P < .001$ for smokers). More visits to physicians increased the likelihood of mammography (OR = 1.11 per additional visit,

TABLE 5—Odds Ratios for Mammography Use in the Last 2 Years for Women Aged 50–75 y (n = 1377)[a]

Independent Variable	Odds Ratio[b]	95% Confidence Interval
Age ≥65	0.81	0.61, 1.06
Married	1.24	0.95, 1.62
Working	1.01	0.75, 1.37
Education	1.05	0.99, 1.11
Smoker	0.59	0.43, 0.79
Breast self-examination	0.92	0.72, 1.16
History	2.22	1.56, 3.16
Physician visits	1.11	1.00, 1.23
No doctor	0.24	0.15, 0.39
Gynecologist visits	2.59	1.71, 3.92
Net price	0.91	0.90, 0.91
Time cost	0.86	0.64, 1.14
Inconvenience	0.53	0.36, 0.78
Income	1.14	1.06, 1.22

[a]Controlling for county of residence.
[b]Odds ratios are calculated for a one-unit change in continuous variables.

P = .047), but education and marital status were not statistically significant when economic variables were included in the model. When economic variables were excluded from the model, married women were more likely to be users (OR = 1.47, P = .003), as were more educated women (OR = 1.08 for each additional year of education, P = .004).

A variable indicating whether a woman's health insurance covered 50% or more of the price of a mammogram was created. This variable and the interactions between it and those variables measuring time cost, inconvenience, and income were added to the model to test the hypothesis that the importance of economic variables may be greater among women with poor insurance coverage. These interactions were not statistically significant, either individually or as a group (likelihood ratio statistic on 3 df = 5.30, P = .151), suggesting that the effects of time cost, inconvenience, and income on mammography use do not vary with insurance coverage. Similarly, to test the hypothesis that price sensitivity is greater among women with low income, interaction effects between net price, time cost, and inconvenience and a dummy variable for women with household incomes above $15 000 were estimated. The interaction effects were not statistically significant, either individually or as a group (likelihood ratio statistic on 3 df = 0.671, P = .880), suggesting that the effects of net price, time cost, and inconvenience are stable across the income groups.

Discussion

As suggested by the theory of demand, the effects of net price and income on the use of mammography among women aged 50–75, residing in four suburban or rural counties in Washington State, were found to be significantly negative and positive, respectively. The analyses provided no evidence that economic variables are more important for low-income or uninsured women. Instead, effects were stable across the subsets of the population investigated.

The effects of economic variables on use of mammography were found to be important, but they do not suggest that improvements in insurance coverage will dramatically increase the use of screening mammography. The data suggest that insurance coverage for mammography was already quite good among women aged 50 to 75 in Washington State in 1989; over half of the women reported that their insurance paid some (usually 80%) or all of the cost of screening mammography. The new legislation can be expected to increase the average proportion paid by insurance from about 60% to about 80%, which can be expected to decrease the average net price from about $32 to about $16. If the coefficients estimated from these cross-sectional data were to be given a causal interpretation, they would suggest that the percentage of women having a mammogram in the last 2 years might increase in response to improved insurance coverage from about 59% to about 61%, a change of less than 4%. Such an increase would be negligible relative to the secular trend in mammography use. This study provides no evidence that the effect of improved insurance coverage would be greater among low-income women.

However, if physicians are influenced in their referral patterns by the improvement in insurance coverage, the legislation could have a greater impact. A woman might depend on her physician to advise her regarding the value of mammography because the physician has better information than she does about the expected medical benefit of alternative medical services. The physician presumably acts in the best interest of the patient, taking into account her ability to pay for mammography as well as his or her assessment of its value to the patient. This phenomenon has long been recognized as the "agency role" of the physician.[16] However, the physician may be ignorant of the individual woman's ability to pay and reluctant to discuss the issue with her. Instead, the physician may assume that insurance will (or will not) pay for screening mammography and treat all or most women accordingly.

To the extent that the legislative mandates change the physician's assumptions regarding insurance coverage for screening mammography, they could influence physicians' recommendations. State legislation mandating coverage of screening mammography by private insurers and federal legislation mandating such coverage by Medicare may have an important impact on use by creating a "standard of care" recognized by all physicians. The strength of the influence of physicians is evident from the statistical significance of the physician variables in the demand equation. If all physicians were to refer women for mammography at the rates reported for the gynecologists, the estimates obtained here suggest that use of mammography might rise from 59% to about 75%.

Mandating insurance coverage for screening mammography will not compensate for lack of coverage for preventive office visits. As has been shown in previous studies, exposure to physicians, especially to gynecologists, is an important determinant of whether breast cancer screening is done, including clinical breast examination as well as mammography use.[19] In this sample of primarily upper-middle-class White women, 19.1% of the women who had not had mammography in the last 2 years reported seeing no physician regularly, compared with 3.4% of those who had had mammography. Of the nonusers, 10.4% reported having no health insurance at all, compared with 3.5% among the users.

Because variables other than economic variables are important in determining the use of mammography, interventions will still be needed to reach women who do not obtain screening even when it is available at low cost. Among the women in this sample, family history of breast cancer was an important predictor of use, but age was negatively related to use. A family history of breast cancer may contribute to awareness and fear of the disease in a way that aging does not. In addition, women may be more aware of family history than aging as a risk factor. Also notable in this study is the high rate of smoking among the nonusers, 26.7%

compared with 15.0% among the users. In the multivariate analyses, smokers were found to be about half as likely to have had mammography as nonsmokers. This result suggests that compliance with physicians' recommendations for mammography may be an important problem among smokers, perhaps because the procedure is generally performed at a different facility at a later time, providing ample opportunity for noncompliance.

Despite suggestions in the literature that time costs are substantial,[20] time cost was not statistically significant in the demand analysis. However, women who preferred to obtain mammography during evenings or weekends rather than weekdays were significantly less likely to have had mammography during the previous 2 years, suggesting that logistics are indeed important, as has been suggested previously.[21] The relative unimportance of time cost in these analyses may be attributable in part to limitations in its measurement: although total time to obtain a mammogram was measured, the value of time to the woman was not measured.

Other limitations of this study include the cross-sectional, self-report nature of the data and the lack of measures of price and insurance coverage for physicians' services. Most mammography screening is the result of a referral after a visit to a primary-care physician for routine care. The cost of preventive services may remain a barrier to use of breast cancer screening when the cost of mammography services is reduced by insurance. A minor limitation is that, because a direct measure of general health status was not available, the effect of the number of visits to physicians, intended to measure exposure to physicians, may be biased. Similarly, household income was not adjusted for family size because the latter was not measured.

A potential limitation of the study is that women may not be well informed regarding price and insurance coverage for screening mammography. Of interest in this regard is that women who guessed the price differed from women who knew the price by only $2 on average in their estimate of mammography price. Similarly, all but 27 of the 161 members of Group Health Cooperative of Puget Sound who were included in the analyses (10.5% of the sample) knew that their insurance covered 100% of the cost of screening mammography. The Group Health Cooperative is a large health maintenance organization that has had a risk-based breast cancer screening program since 1984.[22] There is no out-of-pocket charge for screening mammography for enrollees in the Group Health Cooperative. Of the 27 patients in this group, 18 did not know whether the Group Health Cooperative covered screening mammography, 5 thought that it did not, and the remaining 4 thought that it covered something less than 100% of the cost. External validation by means of a survey of mammography facilities in the four study communities also confirmed that women were quite knowledgeable. On average, the women reported a price of $77, and the mammographers reported a charge of $63 for screening mammography and $94 for diagnostic mammography. A related concern is that the dummy variables indicating health insurer that were used to predict the proportion of the price of screening mammography paid by insurance are crude measures. As a group in the prediction equation, they explained 24% of the variance, suggesting that they served adequately as instrumental variables.

This study is an attempt to carefully measure mammography price and insurance coverage and to analyze the use of mammography within the conceptual framework of the theory of demand. It provides evidence that cost is indeed important as a barrier to use of mammography. However, it should not be assumed that mandatory coverage of screening mammography will solve the problem of underutilization of breast cancer screening among all women aged 50 to 75. Uninsured women, smokers, and women who do not visit a physician regularly may still have relatively low rates of use. The analyses corroborate the abundant evidence in the literature that the key to use of mammography among women aged 50 to 75 is referral by a physician. □

Acknowledgments

This research was supported by grant CA-34847 from the Division of Cancer Prevention and Control of the National Cancer Institute, National Institutes of Health.

The authors acknowledge helpful comments from Douglas Conrad and other scientists at the Cancer Prevention Research Program of the Fred Hutchinson Cancer Research Center.

References

1. Kessler LG, Feuer EJ, Brown ML. Projections of the breast cancer burden to U.S. women: 1990–2000. *Prev Med.* 1991;20: 170–182.
2. Shapiro S, Venet W, Strax P, et al. Ten to fourteen year effect of screening on breast cancer mortality. *J Natl Cancer Inst.* 1982; 69:349–355.
3. Tabar L, Fagerberg G, Duffy SV, Day NE. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health.* 1989;43:107–114.
4. National Cancer Institute Breast Cancer Screening Consortium. Screening mammography: a missed clinical opportunity? *JAMA.* 1990;264:54–58.
5. Zapka JG, Stoddard AM, Costanza ME, Greene HL. Breast cancer screening by mammography: utilization and associated factors. *Am J Public Health.* 1989;79:1499–1502.
6. Lerman C, Rimer B, Trock B, Balshem A, Engstrom P. Factors associated with repeat adherence to breast cancer screening. *Prev Med.* 1990;19:279–290.
7. Rimer BK, Keintz MK, Kessler HB, et al. Why women resist screening mammography: patient-related barriers. *Radiology.* 1989;172:243–246.
8. Kruse J, Phillips DM. F[illegible]rs influencing women's decision to undergo mammography. *Obstet Gynecol.* 1987;70:744–748.
9. Johnson RA, Murata PJ. Demographic, clinical, and financial factors relating to the completion rate of screening mammography. *Cancer Detect Prev.* 1988;11:259–266.
10. O'Leary TJ, deParedes ES, Tritschler J, Barr M. Utilization of screening mammography by primary care physicians. *Appl Radiol.* 1989;18(12):28–33.
11. Thompson GB, Kessler LG, Boss LP. Breast cancer screening legislation in the United States. *Am J Public Health.* 1989; 79:1541–1543. Commentary.
12. Larrakas PJ. *Telephone Survey Methods: Sampling, Selection, and Supervision.* Newbury Park, Calif: SAGE Publications Inc; 1987.
13. King ES, Rimer BK, Trock B, Balshem A, Engstrom P. How valid are mammography self-reports? *Am J Public Health.* 1990;80: 1386–1388.
14. Howard J. Using mammography for cancer control: an unrealized potential. *CA Cancer J Clin.* 1987;37:33–48.
15. Grossman M. On the concept of health capital and the demand for health. *J Political Economy.* March–April 1972;80:223–255.
16. Dionne G, Contandriopoulos AP. Doctors and their workshops: a review article. *J Health Econ.* 1985;4:21–33.
17. *EGRET User's Manual.* Seattle, Wash: Statistics and Epidemiology Research Corporation and Cytel Software Corporation; 1989.
18. Maddala GS. *Econometrics.* New York: McGraw-Hill Inc; 1977.
19. Burg MA, Lane DS, Polednak AP. Age group differences in the use of breast cancer screening tests. The effects of health care utilization and socioeconomic variables. *J Aging Health.* 1990;2:514–530.
20. Hurley SF, Livingston PM. Personal costs incurred by women attending a mammographic screening programme. *Med J Aust.* 1991;154:132–134.
21. Burack RC, Liang J. The acceptance and completion of mammography by older Black women. *Am J Public Health.* 1989; 79:721–726.
22. Taplin SH, Thompson RS, Schnitzer F, Anderman CA, Immanuel V. Revisions in the risk-based breast cancer screening program at Group Health Cooperative. *Cancer.* 1990;66:812–818.

Using Census and Mortality Data to Target Small Areas for Breast, Colorectal, and Cervical Cancer Screening

ABSTRACT

Objectives. The goal of this study was to develop and validate quantitative models for estimating cancer incidence in small areas.

Methods. The outcome for each cancer site was the incidence of disease that had reached a late stage at the time of diagnosis. Two sets of predictors were used: (1) census-based demographic variables and (2) census-based demographic variables together with the cancer-specific mortality rate.

Results. The best models accounted for a substantial percentage of between-area variability in late-stage rates for cancer of the breast (46%), cervix (61%), and colon/rectum (58%). A validation procedure indicated that correct identification of small areas with high rates of late-stage disease was two to three times more likely when model-based estimates were used than when areas were selected at random.

Conclusions. Additional testing is needed to establish the generality of the geographic targeting methodology developed in this paper. However, there are strong indications that small-area estimation models will be useful in many regions where planners wish to target cancer screening programs on a geographic basis. (*Am J Public Health.* 1994;84:56–61)

Howard F. Andrews, PhD, Jon F. Kerner, PhD, Ann G. Zauber, PhD, Jeanne Mandelblatt, MD, MPH, John Pittman, MS, and Elmer Struening, PhD

Introduction

Models for cancer screening programs typically focus on subgroups of the general population on the basis of age- and sex-specific incidence, prevalence, and mortality rates of a particular cancer. Although analyses of individual characteristics indicate who should be screened, small-area analysis can be used to determine where to set up screening programs so that high-risk populations can be more easily recruited. In previous studies of small-area variation in cancer incidence, we quantified the benefits of geographically targeting cancer screening programs, using clustered census tracts as the unit of analysis.[1,2] Those analyses, based on cancer registry data, indicated that the efficiency of breast cancer screening programs can be substantially improved by targeting communities with known high incidence rates.

In many localities, there is no cancer registry to provide incidence data at the community level, so targeting based on known small-area incidence, using the methodology described above, is not possible. Here we describe a two-step technique for making geographic targeting available to these localities:

1. Develop a small-area multiple regression model relating incidence for a particular cancer site to mortality, census-based variables, or both in a locality where cancer registry data are available.
2. Apply the model in the locality of interest, where cancer registry data are not available, and obtain an estimated incidence value for each small area.

The outcome in the regression models is the incidence rate of disease staged as regional or distant at the time of diagnosis (late-stage incidence). The main goal of screening programs is to reduce this rate by identifying in situ and localized cancer in individuals who would otherwise have developed late-stage disease before diagnosis and treatment.

Methods

Geocoding Cancer Incidence and Mortality

New York City cancer incidence data for the years 1976 through 1982 were obtained from the New York State Department of Health Cancer Registry. Mortality data for the years 1977 through 1983 were obtained from the New York City Department of Health. The geographic coding of these data has been previously described.[1] Denominators for small area rates were obtained from 1980 US census data (Summary Tape File 3a).

Incidence, mortality, and census data were aggregated to health areas, administrative units used by the New York City Department of Health. Each health area, with an average population of about 21 000, consists of four to six census tracts. Data were obtained for four of the five boroughs that make up New York City: Bronx, Brooklyn, Manhattan, and

Howard F. Andrews, John Pittman, and Elmer Struening are with the New York State Psychiatric Institute, New York, NY. At the time of the study, Jon F. Kerner, Ann G. Zauber, and Jeanne Mandelblatt were with the Memorial Sloan-Kettering Cancer Center, New York, NY. Dr Kerner is now with the Vince T. Lombardi Cancer Research Center, Georgetown University Medical Center, Washington, DC.

Requests for reprints should be sent to Howard F. Andrews, PhD, Epidemiology Department, New York State Psychiatric Institute, 722 W 168th St, Box 47, New York, NY 10032.

This paper was accepted May 11, 1993.

Queens. To reduce rate instability due to small denominators, health areas with a total 1980 population of less than 5000 were excluded from the analysis. Areas in which more than 10% of the total population resided in an institution were also excluded. The 22 health areas excluded by these criteria contained 4.2% of the total population of the four boroughs. The 319 health areas that were included contained a total population of 6 637 686 in 1980.

Rate Calculation

It is customary to adjust for age when comparing geographic variability in cancer rates. However, when targeting geographic areas for screening service delivery, a concentration of age groups at higher risk for a particular cancer should be considered as part of the planning process. We therefore calculated crude age-truncated rates for each area by dividing the total number of incident late-stage cases at or above a specific age by the total area population of the appropriate sex at or above the same age. For cervical cancer, we used the incidence of invasive disease as the measure of late stage. For breast and colorectal cancer, late stage was defined according to the Cancer Registry categories of regional or distant disease.

For each cancer site, an age-truncated crude late-stage incidence rate and an age-truncated crude mortality rate were calculated. For breast cancer, incidence and mortality rates were calculated for women aged 35 years and older, for cervical cancer, rates were calculated for women aged 20 years and older; for colorectal cancer, we combined men and women aged 45 years and older in calculating rates. No upper age limit was set for any site. These age-truncated rates include more than 95% of registry cases for each of the three sites. The total number of late-stage cases used for rate calculation was 11 634 for breast cancer, 16 140 for colorectal cancer, and 3640 for cervical cancer. For mortality rate calculation, the numerator for each cancer site represents all individuals for whom the primary cause of death was cancer of that site during the period 1977 through 1983: 9132 for breast cancer, 9212 for colorectal cancer, and 1285 for cervical cancer. Incidence data and mortality data were not linked: the numerator for each mortality rate includes all deaths for that cancer site in the stated time period, not just deaths of individuals reported to the Cancer Registry during the period for which incidence rates were calculated.

All rates are expressed in terms of yearly averages per 100 000 population. The formula for rate calculation is

$$[(N/7)/T] \cdot 100\,000,$$

where N is the total number of late-stage cases in a given health area for a particular cancer site among individuals in the designated age group during the 7-year study period and T is the total 1980 population in the health area for the site-defined age and sex group. Hereinafter, age-truncated late-stage incidence rates for breast and colorectal cancer and the rate of invasive cervical cancer will be referred to as late-stage rates.

Selection of Predictors for Use in Models to Estimate Late-Stage Incidence Rates

In choosing census-based predictors, we were guided by two criteria: known etiologic relationship to one or more of the three cancers under consideration and absence of multicollinearity among predictors. We were aided in variable selection by previous research that identified dimensions of demography and housing characteristics in New York City.[3] Beginning with the 14 census-based variables that defined these dimensions, we used multiple regression analysis to isolate a subset of four variables that met the above criteria and that, when used together, accounted for nearly as much variability (adjusted R^2) in each of the late-stage rates as the full set of 14 variables: (1) percentage of the population aged 65 years and older; (2) percentage of households with income higher than $50 000 per year; (3) percentage of the population (aged 15 years and older) divorced or separated; and (4) percentage of women in the labor force with one or more children aged 16 years or younger. Descriptive statistics for these variables, together with mortality rates and late-stage incidence rates, are shown in Table 1. For analytic purposes, we applied log or square root transformations to skewed variables to approximate a normal distribution, as indicated in Table 1.

Data from all four boroughs were used to develop a multiple regression model for each of the three late-stage rates, with the four census variables and the appropriate mortality rate as predictors. To simulate a situation in which targeting is to be done in a locality for which cause-specific mortality data are not available or not accessible, we also developed a model for each cancer in which only census variables were used as predictors. The addition of interaction terms did not improve the adjusted R^2 values for any of the regression models.

TABLE 1—Descriptive Statistics for Outcome and Predictor Variables (n = 319 Health Areas)

	Mean	SD
Outcome variables—incidence rates[a]		
Late-stage breast cancer	91.72	23.92
Invasive cervical cancer[b]	22.44	12.76
Late-stage colorectal cancer	94.48	24.58
Predictors—mortality rates[a]		
Breast cancer[b]	71.66	21.85
Cervical cancer[b]	8.22	6.68
Colorectal cancer	79.22	21.32
Predictors—census variables		
% persons aged 15+ y divorced or separated	11.98	4.81
% working mothers with children <18 y	29.48	10.43
% households with income above $50 000[c]	3.30	4.76
% persons aged 65 y and older	13.16	5.70

[a]Expressed as cases per 100 000 population.
[b]A square root transformation was applied to this variable in the regression models.
[c]A log transformation was applied to this variable in the regression models.

Model Validation

For each cancer site, we used the following procedure to determine the accuracy with which a model developed in one geographical area could predict late-stage rates in another area.

1. One of the four boroughs of New York City—for example, Queens—was designated as the geographic area for which estimated rates of late-stage disease for a particular site—say breast cancer—were needed.

2. A multiple regression model for estimating late-stage breast cancer rates was developed, using only data from the other three boroughs—in this case, Bronx, Manhattan, and Brooklyn.

3. The regression model was used to calculate an estimated late-stage rate for each health area in Queens. This was done by taking the known value of each predictor for a particular health area in Queens and multiplying by the appropriate unstandardized regression coefficient specified in the three-borough model. These

TABLE 2—Correlation[a] between Estimated and Actual Late-Stage Incidence Rates, Based on Models with Mortality and Census Variables and Census Variables Only

Borough	Census and Mortality	Census Only
Breast cancer		
Brooklyn	.69	.62
Bronx	.74	.70
Manhattan	.63	.46
Queens	.53	.40
Mean	.85	.55
Cervical cancer		
Brooklyn	.78	.78
Bronx	.73	.63
Manhattan	.79	.72
Queens	.74	.67
Mean	.76	.70
Colorectal cancer		
Brooklyn	.77	.60
Bronx	.80	.83
Manhattan	.68	.44
Queens	.68	.50
Mean	.73	.59

[a]Pearson's product-moment correlation coefficient (*r*).

terms were then summed and the value of the model intercept was added, yielding the estimated late-stage rate for that health area. This process was repeated for each health area in Queens.

4. The model was validated by correlating the estimated late-stage rates for the 74 health areas of Queens, based on the model developed with data from the other three boroughs, with the actual late-stage rates in those same areas, based on data from the New York State Department of Health Cancer Registry.

For each cancer site, the above steps were repeated. Data from each unique combination of three boroughs were used to build the model and obtain an estimated rate of late-stage disease for each health area in the excluded borough. For each site and borough, one model was built with the four census-based variables and the site-specific mortality rate as predictors; the other model included only the four census-based predictors.

Results

The correlation between actual late-stage rates and late-stage rates estimated from the appropriate model is shown for each combination of borough and cancer site in Table 2. In general, the association between estimated and actual rates was highest for invasive cervical cancer, for which all correlations were greater than .7 when mortality and census data were used in the estimation model. As shown in the table, all correlations were greater than .6 even when census data alone were used. The lowest correlations were obtained for the breast cancer models: the mean correlation was .65 when mortality was included and .55 when only census variables were used. For colorectal cancer, the correlation between estimated and actual late-stage rates varied from a low of .68 in Manhattan to a high of .80 in the Bronx when mortality and census data were used. When only census data were used in the model, correlations ranged from .44 in Manhattan to .83 in the Bronx.

From a planner's perspective the key question with respect to model validity is, What is the relationship between actual and estimated rates *in the areas targeted to receive a particular service?* To answer this question, we follow a hypothetical scenario in which planners have decided to target for screening the health areas with the highest estimated late-stage rates, defined as those areas in the top quintile (20%).

TABLE 3—Accuracy of Targeting on the Basis of Estimation Models

	Model Type			
	Census and Mortality		Census Only	
Borough	Top 20%	Top 40%	Top 20%	Top 40%
Breast cancer				
Brooklyn	45	81	32	64
Bronx	73	82	73	82
Manhattan	47	74	40	67
Queens	47	73	40	53
Mean	53	76	46	67
Cervical cancer				
Brooklyn	59	91	41	82
Bronx	64	82	55	82
Manhattan	73	100	53	93
Queens	53	80	53	86
Mean	62	88	51	86
Colorectal cancer				
Brooklyn	68	82	50	77
Bronx	64	100	82	91
Manhattan	67	87	53	60
Queens	60	73	53	80
Mean	65	86	60	77

Note. Numbers in table are percentages of targeted health areas (areas with estimated late-stage rates in the top 20%) that fell into the top 20% or top 40% of all health areas with respect to actual late-stage incidence rates. Higher percentages indicate greater accuracy.

To determine the accuracy of this targeting process, we calculated the percentage of targeted areas in each borough that fell in (1) the upper 20% of health areas with respect to *actual* late-stage rates and (2) the upper 40% of health areas with respect to *actual* late-stage rates. If the model provided perfect predictions, 100% of the targeted areas would be in the upper 20% with respect to actual late-stage rates. In the worst-case scenario, in which the model had no predictive power, one would expect that on average only 20% of the targeted areas would fall in the upper quintile of actual late-stage rates—the proportion expected by chance. Similarly, with perfect prediction, 100% of the targeted areas would be in the upper 40% of areas with respect to actual late-stage rates; in the worst case, only 40% of the targeted areas would be in the upper two quintiles of actual late-stage rates. The normal approximation to the binomial distribution was used to determine whether the actual number of targeted areas in the upper quintile was significantly greater than the number expected by chance. The results are shown in Table 3.

The concordance between targeted areas and the upper quintile of actual late-stage rates is greatest for invasive cervical and colorectal cancer and lowest for breast cancer. For all cancer sites, the concordance between upper quintile of estimated and actual rates is considerably better than what would be expected by chance ($P < .00001$). For example, as indicated in the first row, first column of Table 3, 45% of the areas in the upper quintile of estimated late-stage rates of breast cancer in Brooklyn (the targeted areas) were also in the upper quintile of actual late-stage rates—more than twice the chance expectation of 20%. As indicated in the first row, second column, 81% of the health areas that were in the upper quintile with respect to estimated late-stage rates were either in the highest or second highest quintile in terms of actual late-stage rates, more than twice the 40% rate of concordance expected by chance. Consistent with the correlational analyses (Table 2), concordance between estimated and actual rates is better in models using mortality and census-based predictors than in models using only census-based predictors.

Maps of the results with respect to one of the outcomes, invasive cervical cancer, are shown for the Bronx (Figure 1), providing visual confirmation of the strong relationship between estimated and actual rates. The maps also show that the high-rate areas that would be targeted for screening in the Bronx tend to cluster geographically. Such clustering should enhance the efficiency with which screening resources can be geographically allocated.

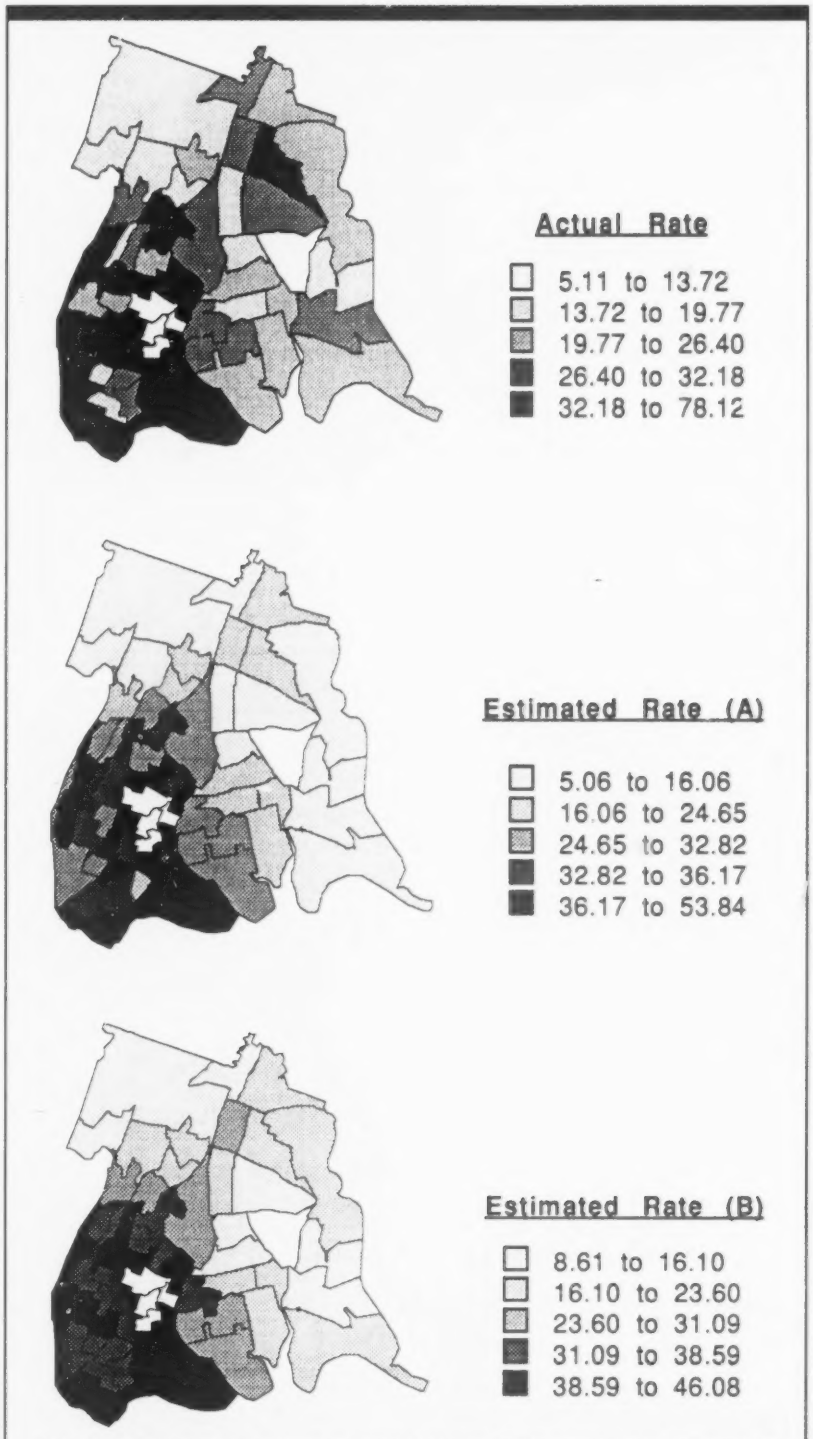
"Best" Models

In the validation process described above, several models were developed for each cancer site, each based on data from three boroughs. Given the apparent viability of targeting small areas on the basis of census data alone or census and mortality data, the question arises, What specific model should be used in applying this approach in a city or region in which cancer registry data are not available?

To develop a set of "best" models for use by planners, we used data from all four boroughs to maximize the stability of parameter estimates. As shown in Table 4, the four census variables alone explained a statistically significant proportion of variability in each of the late-stage rates ($P < .001$): the adjusted R^2 was .33 for breast cancer, .54 for invasive cervical cancer, and .39 for colorectal cancer. In the models using breast cancer and colorectal cancer as outcomes, only one census variable accounted for a significant proportion of variability in late-stage rates—the percentage of the population in the health area that was 65 and older. For cervical cancer, two census variables were significantly associated with the rate of invasive disease: (1) percentage of the population divorced or separated and (2) income level—poor areas had significantly higher rates than wealthier areas.

For each cancer site, the addition of the site-specific mortality rate to the model added significantly to the proportion of explained variability. The greatest increment in adjusted R^2 resulting from the addition of mortality rate as a predictor occurred for colorectal cancer (19%). Moreover, for this cancer site two census variables that were not significant in the census-only model, percentage of working mothers and percentage of population divorced or separated, did make significant contributions to R^2 when the mortality rate was added.

When the site-specific mortality rate was used without census variables in the model, a substantial proportion of variability in late-stage rates was accounted



Note. Areas with rates in the upper quintile (20%) of all health areas—those that would be targeted in a screening program for cervical cancer—are shown in the darkest shading. Unshaded (white) areas had populations that did not meet criteria for inclusion in the analysis. Estimated rates, based on models developed using data from health areas in Brooklyn, Manhattan, and Queens, are shown in the two lower maps. In model A, estimated rates were based on regression analysis in which both mortality rates and census variables were used as predictors. In model B, rates were based on an analysis in which only census variables were used.

FIGURE 1—Geographical distribution of actual and estimated rates of invasive cervical cancer in health areas in the Bronx (n = 57).

TABLE 4—Multiple Regression Models for Estimating Late-Stage Incidence

Model Type	Regression Coefficient (Unstandardized) Mortality	Divorced/ Separated	Working Mother	High Income	Elderly	Intercept	R^2
Breast cancer							
Census and mortality	7.01	...	...	...	10.46	−3.79	.46
Census only		...	...	...	12.47	30.72	.33
Cervical cancer							
Census and mortality	.39	.09	...	−.52	...	3.00	.61
Census only		.13	...	−.67	...	3.75	.54
Colorectal cancer							
Census and mortality	.60	−.58	.30	...	10.86	6.57	.58
Census only		...	...	...	19.33	26.07	.39

Note. The regression model accounts for a significant proportion of the variability (R^2) in late-stage incidence rate in all models ($P < .001$). R^2 values are adjusted for number of variables in model. Coefficients of all variables shown have a significant relationship to the outcome variable in the model ($P < .05$). Ellipsis points indicate a variable excluded from the final model because its effect when included in a preliminary analysis was nonsignificant.

TABLE 5—Multiple Regression Models for Estimating Late-Stage Incidence Based Only on Mortality

Cancer	Mortality Regression Coefficient	Intercept	R^2
Breast	10.03	8.07	.37
Cervical	.75	2.58	.41
Colorectal	.81	30.24	.49

Note. The regression model accounts for a significant proportion of the variability (R^2) in late-stage incidence rate in all models ($P < .001$). R^2 values are adjusted for number of variables in model.

for: 37%, 41%, and 49% for cancers of the breast, cervix, and colon/rectum, respectively (Table 5). Thus, site-specific mortality data could be used without census data to target areas where late-stage disease is high. However, since census data are available everywhere in the United States, and since the addition of census-based variables adds 10% to 20% to the explained variability in the models, there should be no need to develop or implement mortality-only models. The only situation in which such models might be valuable is one in which (1) there has been rapid change in the demographic characteristics of the neighborhoods under consideration for community-based screening programs and (2) available census data were collected too long ago to capture these changes.

To test the possibility that there was a spurious inflation of R^2 in the models owing to the use of census-based population counts in the calculation of both predictors and outcome rates, we entered the factor $1/N$ in each equation, where N was the same value used in the denominator of the age-truncated rate. The results indicated that there was no spurious correlation: in each model the significance of each predictor was unaffected by the addition of $1/N$, and the values of the regression coefficients remained approximately the same as those shown in Table 4. However, for two of the cancer sites, breast and cervix, the term $1/N$ was itself significant in the model. In the case of late-stage breast cancer, the adjusted R^2 was increased by 2% by the addition of $1/N$ to the model. For both sites, the sign of the regression coefficient for $1/N$ was positive, indicating that when the other variables in the model were held constant in the regression model, late-stage rates tended to be higher in areas of low population. This suggests that low population, or some factor associated with it, makes an independent contribution to the risk of late-stage disease.

Are the Models Applicable to Other Regions?

One test of the potential generality of the models developed in this paper is whether model-based estimates are accurate in geographic areas that differ in terms of demography, outcome rates, or both from the geographic areas in which the models were developed.

To determine the extent of differences among the four boroughs used in the model validation process, we performed a series of one-way analyses of variance to compare mean values of site-specific mortality late-stage rates and each of the census-based predictors in the four boroughs of New York City, using health area as the unit of analysis. The results indicate that the four boroughs differed substantially in terms of key demographic and outcome characteristics. All pairs of boroughs except Manhattan and Brooklyn differed significantly in terms of mean values of at least one late-stage rate. Four of the six pairs differed in terms of at least one site-specific mortality rate, and all six pairs differed in terms of mean value of at least one demographic characteristic.

Discussion

The results indicate that a quantitative model developed in one geographically defined area on the basis of available census data can be used to obtain good estimates of late-stage rates for breast, colon/rectum, and cervical cancer in another geographic area. The accuracy of estimation is improved for all three cancer sites when the site-specific mortality rate is included as a predictor. Because disease that has progressed to a late stage at the time of detection is likely to result in death, it is not surprising that mortality data considerably enhance the power of the model to estimate late-stage rates. More unexpected is the finding that without mortality rate as a predictor, and using only four variables derived from census data, the models can account for a substantial proportion of geographic variability in late-stage rates of all three cancer sites. Therefore, even in localities without ready access to population-based mortality data, it should be possible to estimate late-stage rates accurately and to successfully target screening programs to high-risk communities. However, all of the areas used in developing and testing of the models presented in this paper are located in New York City. Ultimately, these models will be useful only if they can be applied in targeting community-based screening programs in other regions of the United States and the world. Thorough testing will require comparison of model-based estimates with actual rates in other regions of the country where cancer registry data are available.

There are a number of indications that these models will work well in other geographical regions. First, the results demonstrate that the four boroughs of New York City differ substantially along

the dimensions of demography and cancer outcome examined in this paper. This indicates that the models are robust with respect to demographic and disease characteristics of the region for which estimates of late-stage cancer are required.

An examination of the predictors used to estimate late-stage rates also suggests that the models will be widely applicable. Site-specific mortality rates would be expected to be strongly associated with late-stage incidence in any community. For breast and colorectal cancer, an increase in overall incidence as well as late-stage disease with advancing age has been well documented,[4,5] so the percentage of an area's population that is 65 years old or older would be expected to predict late-stage rates for these two sites in any locality. Similarly, invasive cervical cancer and late-stage colorectal cancer are known to be more prevalent in poor populations,[4,6] so the association of low income and socioeconomic factors with late-stage incidence that is quantified in the models developed in New York City should apply in other areas as well.

However, the relationship between predictors and cancer outcomes will no doubt be attenuated if the models are applied to regions in which there is little variability, as would be the case in, say, a region made up of middle-income suburban communities. The models developed in this paper are therefore most appropriate in areas that contain substantial demographic variability among the communities under consideration for targeting.

The finding that performance of the models was poorest for breast cancer is consistent with the results of a previous analysis,[2] which indicated that of the three cancer sites examined, the year-to-year reliability of late-stage rates was lowest for breast cancer. When the reliability of any outcome measure is relatively low, the ability to account for variability in the outcome on the basis of other measures will be limited.[7] However, the results indicate that even for breast cancer, the models correctly identified high-risk geographic areas at considerably better than twice the level that would be expected by chance.

Methods used in some small-area analyses involving health outcomes have been criticized for failing to adequately test the null hypothesis that observed differences between late-stage rates are due to random variation.[8,9] When data from multiple years are available, the method we have developed to quantify year-to-year rate stability[2] and that was applied to the outcomes examined in this paper can be used to directly test the null hypothesis of random between-area variation.

Can the estimation models be improved? Some of the unexplained variability in each model may reflect measurement error or factors that are specific to particular geographic regions. However, it may be possible to improve the R^2 value of the models by including additional demographic factors. The finding that low population is related to high rates of late-stage disease, after controlling for sociodemographic factors, suggests one direction for future research. In New York City, health area boundaries were originally defined in such a way that each area had a total population of approximately 20 000. Health areas with low population have resulted from loss of housing in certain areas.[10] The impact of loss of housing and social disintegration on the health status of individuals who remain in these areas is well documented.[11] It may be that census-based measures of housing loss or other indicators of deterioration over time will improve the ability of estimation models to identify appropriate sites for targeting community-based cancer screening and other health-related intervention programs.

In conclusion, there is good reason to believe that the relationships among late-stage cancer incidence, mortality, and demographic characteristics that have been quantified in the models presented in this paper will apply in many urban and nonurban regions. If this hypothesis can be confirmed in future research, geographically based targeting of cancer screening programs will no longer be limited to areas in which population-based cancer registries have been established. □

Acknowledgments

This investigation was supported in part by NCI grant J1 R01 CA 46586.

We would like to express our thanks to Philip C. Nasca, PhD, Director, Bureau of Cancer Epidemiology, New York State Department of Health, and to William Burnett, MD, Director, New York State Cancer Incidence Registry, who made Registry information available and generously shared their expertise in the analysis of these data. We thank Drs Patricia Cohen, Bruce Link, and Rodrick Wallace and two anonymous reviewers for their valuable comments on earlier versions of the manuscript.

References

1. Kerner JF, Struening E, Pittman J, Andrews H, Sampson N, Strickman N. Small area variation in cancer incidence and mortality: a methodology for targeting cancer control programs. In: Engstrom P, Anderson P, Mortenson L, eds. *Advances in Cancer Control: Epidemiology and Research.* New York, NY: Alan R. Liss Inc; 1984:225–234.
2. Kerner J, Andrews H, Zauber A. Geographically-based cancer control: methods for targeting and evaluating the impact of interventions on defined populations. *J Clin Epidemiol.* 1988;41:543–553.
3. Pittman J, Andrews H, Struening E. The use of zip coded populations in social area studies of service utilization. *Eval Program Plann.* 1988;9:309–317.
4. Mandelblatt J, Andrews H, Kerner J, Zauber A, Burnett W. Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type. *Am J Public Health.* 1991;81:646–649.
5. Ries L, Henkey B, Edwards B, eds. *Cancer Statistics Review.* Bethesda, Md: National Institutes of Health; 1990. Publication 90-2789.
6. Baquet CR, Horm JW, Gibbs T, Greenwald P. Socioeconomic factors and cancer incidence among blacks and whites. *J Natl Cancer Inst.* 1991;83:551–557.
7. Fleiss JL. *The Design and Analysis of Clinical Experiments.* New York, NY: John Wiley & Sons Inc; 1986.
8. Diehr P, Cain K, Connell F, Volinn E. What is too much variation? The null hypothesis in small area analysis. *Health Serv Res.* 1990;24:741–771.
9. Diehr P. A small area simulation approach to determining excess variation in dental procedure rates. *Am J Public Health.* 1990; 80:1343–1348.
10. Struening EL, Wallace R, Moore R. Housing conditions and the quality of children at birth. *Bull N Y Acad Med.* 1990;66:463–478.
11. Wallace RW, Wallace D. Origins of public health collapse in New York City: the dynamics of planned shrinkage, contagious urban decay and social disintegration. *Bull N Y Acad Med.* 1990;66:391–434.

Changes in the Use of Screening Mammography: Evidence from the 1987 and 1990 National Health Interview Surveys

Nancy Breen, PhD, and Larry Kessler, ScD

ABSTRACT

Objectives. **Mammography rates reported by women in the National Health Interview Surveys of 1990 and 1987 are examined. Why this screening modality is not more frequently used is explored.**

Methods. **Data from the 1987 and 1990 National Health Interview Surveys, conducted by the National Center for Health Statistics, are cross-tabulated and compared.**

Results. **In 1987, approximately 17% of women over 40 years of age reported having had a screening mammogram in the previous year. In 1990, the rate doubled. Race declined in importance; income and education remained strong, positive predictors of screening.**

Conclusions. **Despite this dramatic increase, two thirds of women are not having screening mammograms. Use was not higher primarily because women did not realize that screening mammography tests for breast cancer in asymptomatic women. Primary care physicians are the main source of health education for screening mammography. The data suggest that public health programs to promote screening mammography should especially target primary care physicians and women with low incomes and education. Likewise, health care providers should ensure that their patients are referred to facilities that deliver high-quality mammography at low cost to make the procedure more accessible. (*Am J Public Health.* 1994;84:62-67)**

Introduction

Since publication of the results of the New York Health Insurance Plan randomized trial for breast cancer screening in 1982, mammography has gained acceptance in the medical community[1,2] and has increasingly been used for the early detection of breast cancer in asymptomatic women.[3,4] That trial provided the first scientific evidence that mammography and breast physical exams reduce mortality from breast cancer in women 50 years of age and older and possibly in women 40 to 49 years of age.[5-7]

Only with the release of the National Health Interview Survey (NHIS) Public Use Tape in 1990 did trend analysis of the use of mammography by age, race, and socioeconomic status (SES) with a nationally representative sample become possible. Previous researchers had analyzed change over time in use of breast cancer screening modalities based on local, self-selected, or convenience samples.[3,8-10]

Each year, the NHIS includes supplemental questions on different topics of interest to other federal agencies. The National Cancer Institute (NCI) sponsored a cancer control supplement to the 1987 NHIS to assess knowledge of, attitudes toward, and use of a wide variety of cancer screening modalities, including breast physical examination and, for the first time, mammography. The health promotion and disease prevention supplement to the 1990 NHIS included similar questions on breast cancer screening.

In 1987, women were asked about all three widely recommended breast cancer screening modalities: mammography, breast physical examination, and breast self-examination. These and subsequent data consistently showed that age, race, income, and education contribute in predictable ways to whether women are screened for breast cancer.[11-14] Women who have had a mammogram are likely to be younger[3,11,15-17] (although older women are more likely to develop breast cancer[3,11,18]) and White[4,11] and to have a higher income, more education, a usual source of health care,[4,11,16,19,20] and a regular physician (especially a gynecologist or internist).[10] Finally, women enrolled in health maintenance organizations in 1987 were nearly twice as likely to have had a mammogram and 11% more likely to have had a breast physical examination than women relying on other sources of health care.[21]

A recent analysis of 1990 NHIS data used multivariate logistic regression to estimate the likelihood of women 40 to 75 years of age having regular mammograms. Higher income, use of other kinds of screening, having a regular source of health care, living in the West or a metropolitan area, and engaging in healthy practices were all associated with regular mammography.[22]

Our study measured and compared the proportion of White, Black, and Hispanic women who reported being screened in 1987 and 1990. The detailed trend analysis reported here is designed to encourage other researchers to probe further the determinants of use of mammography and breast physical examination. Trends by age, race, usual source of health care, rural or urban location, family income, and educational attainment are examined and their interactions explored.

The authors are with the Applied Research Branch, National Cancer Institute, Bethesda, Md.

Requests for reprints should be sent to Nancy Breen, PhD, Applied Research Branch, National Cancer Institute, EPN 313, Bethesda, MD 20892.

This paper was accepted October 13, 1993.

The National Health Interview Survey

The NHIS is a continuing annual nationwide survey of approximately 49 000 households in the civilian noninstitutionalized population of the United States conducted by the National Center for Health Statistics.[23] Information is gathered on health knowledge, attitudes, and practices for various population groups, including those defined by age, sex, race, family income, geographic region, and place of residence.[24,25]

Blacks are regularly oversampled for the NHIS, but Hispanics are not. The 1987 cancer control supplement oversampled the Hispanic population in order to obtain more accurate estimates for that group. Data presented here come from this supplement, administered to half of the households in the survey, and the health promotion and disease prevention supplement of 1990, administered to all households. Only women 40 years of age and older (6771 respondents in 1987 and 12 808 in 1990) were asked questions on breast cancer screening.

Data from the two surveys are comparable except for minor changes in the questions.[24,25] The series of questions in 1987 began by defining a mammography and then asking (as screening question) respondents whether they had ever heard of mammography. If not, respondents were not asked whether they had ever had one. This may have reduced the response rate slightly in 1987 relative to 1990, when the screening question was not included; however, other studies show similar recent dramatic increases in rates of screening mammography.[3,26,27] Mammography questions were also asked in the 1992 and 1993 NHIS and will be included in the 1994 version.

Methods

The SUDAAN program was used for this analysis. Sample design information was used to calculate population estimates for the 1987 and 1990 NHIS, including proportions and design-appropriate standard errors and tests of significance.[28]

The SUDAAN logistic regression procedure was used to test the hypothesis of group differences in a regression model; screening behavior in the previous year was the dependent variable, and race or ethnicity, income, education, or an interaction was the predictor variable (at the conventional $P < .05$ significance level).

TABLE 1—Percentage of Women Reporting Screening Mammograms in the Last Year: 1987 and 1990

Characteristic	1987 No. of Respondents	1987 Mammogram, %	1987 SE	1990 No. of Respondents	1990 Mammogram, %	1990 SE
Age, y						
40–44	1 035	15	1	2 134	31	1
45–49	737	20	2	1 605	33	1
50–54	677	21	2	1 320	39	2
55–59	762	19	2	1 303	39	2
60–64	776	18	2	1 420	37	2
65–69	827	17	1	1 489	34	2
70–74	701	15	1	1 321	30	1
75–79	608	14	2	1 085	28	2
80+	607	8	1	1 183	18	1
Total	6 730	17	1	12 860	33	1
		$P < .01$			$P < .01$	
Race/ethnicity						
White	5 406	18	1	10 255	33	1
Black	883	14	1	1 727	32	1
Hispanic	354	13	2	664	31	2
		$P < .03$			$P < .5$	
Educational attainment, y						
Less than 12	2 308	10	1	3 777	23	1
12	2 583	18	1	5 014	33	1
More than 12	1 809	23	1	4 028	41	1
		$P < .01$			$P > .01$	
Family income						
Less than $20 000	3 681	11	1	6 151	24	1
At least $20 000	2 882	22	1	6 376	39	1
		$P < .0000$			$P < .01$	
Area of residence						
Urban	4 957	19	1	9 606	35	1
Rural	1 173	12	1	3 254	26	1
		$P < .01$			$P < .01$	

The hypothesis of differences between years was tested with 95% confidence intervals.

Results

The proportion of women over 40 years old who used mammography almost doubled between 1987 and 1990 (Table 1), from 17% to 33%. (Approximately 3% had nonscreening mammograms in both years.) Differences reported are statistically significant unless otherwise indicated. In both years, rates of use peaked in the sixth decade of life and declined thereafter.

Rates of screening mammography among women who reported no usual source of health care more than doubled between 1987 and 1990, from 6% to 15%. They also doubled for women with a usual source of health care, from 18% to 36%.

Mammogram Use among Blacks, Whites, and Hispanics

Rates of mammography screening most closely reflect the behavior of White women because they constituted 79% of all respondents in 1987 and 80% in 1990. Eighteen percent reported being screened in 1987 and 33% in 1990.

Black and Hispanic women reported having a screening mammography in the previous year less often than White women in 1987. Between 1987 and 1990, however, their rates of use increased more rapidly than did those of White women so that, in 1990, all three racial-ethnic groups reported screening mammography at nearly the same rate. Proportional differences by race and ethnic group were statistically significant in 1987 but not in 1990.

Use of Screening by Income, Education, and Location

Table 1 shows use of screening mammograms by three educational levels: less than 12 years of school completed, 12 years completed, and more than 12 years completed. Although greater use was associated with more education in both surveys, the gap diminished relatively (but not absolutely) between 1987 and 1990.

TABLE 2—Proportion of Women Reporting a Screening Mammogram, by Race, Family Income, and Educational Attainment: 1987 and 1990

	1987			1990		
	No. of Respondents	Mammogram, %	SE	No. of Respondents	Mammogram, %	SE
Family income						
Less than $20 000[a]						
White	2 795	12	1	4 564	24	1
Black	642	11	1	1 143	24	2
Hispanic	205	10	2	352	24	2
Total	3 681	11	1	6 151	24	1
At least $20 000[b]						
White	2 471	23	1	5 421	39	1
Black	225	21	3	544	44	2
Hispanic	140	16	4	293	39	3
Total	2 883	22	1	6 376	39	1
Educational attainment, y						
Less than 12[c]						
White	1 638	11	1	2 602	23	1
Black	453	9	1	762	24	2
Hispanic	177	10	3	337	23	3
Total	2 308	10	1	3 777	23	1
12						
White	2 198	19	1	4 216	33	1
Black	260	19	3	571	32	2
Hispanic	103	14	4	170	37	4
Total	2 583	18	1	5 014	33	1
More than 12						
White	1 547	24	1	3 415	41	1
Black	167	18	4	384	46	3
Hispanic	70	18	5	151	43	6
Total	1 809	23	1	4 028	41	1

[a]For 1990, $P < .01$.
[b]For 1987, $P < .01$.
[c]For 1987 and 1990, $P < .01$.

A single income cutpoint of $20 000 was used in this analysis, although more detailed income data show similar patterns.[13] Twice as many women whose family incomes were $20 000 or more in 1987 reported a screening mammogram as those whose family incomes were lower. Since mammography use more than doubled for women in lower income families and did not double for women in higher income families, the gap in screening rates by income declined. However, it remained statistically significant.

In both 1987 and 1990, women living in metropolitan areas reported more frequent recent screening with mammography than their nonmetro counterparts, even after age, race, income, and education were controlled. This is not surprising since facilities tend to be concentrated in high-income urban areas.[29]

Income, Education, and Race

Because they are correlated with race and ethnicity in the United States, income and education for women in the three predominant race and ethnic groups were examined and compared. Table 2 shows the proportion of White, Black, and Hispanic women more than 40 years of age who reported a screening mammogram by family income and educational attainment.

Women in families with annual incomes of $20 000 or more in 1987 reported using mammography at higher rates than women in families with lower incomes. Between 1987 and 1990, screening mammography rates rose 23 percentage points each for Black and Hispanic women and 16 percentage points for White women. Among White, Black, and Hispanic women living in households with incomes of less than $20 000, rates rose 12%, 13%, and 14%, respectively, so that by 1990 all three groups had risen to 24%. Differences by race and ethnicity were not statistically significant in either year after income and education were controlled.

Hispanic women reported using mammography screening at the lowest rates of any of the race and ethnic groups under study, yet mammography use increased most among Hispanic women between 1987 and 1990.

The greatest increase in use of screening mammography associated with education occurred among Black women with less than a high school education. Among Hispanic women, the greatest gains were for those with a high school education. Among White women, gains were greater among those who were less educated and declined with higher educational attainment.

Table 3 shows women who reported a screening mammogram, by age group, race, and percentage of these who graduated from high school. Older Black women showed the greatest gain between 1987 and 1990. The decline in mammography use associated with age in 1987 was far less apparent among Black women in 1990, except for those 80 years of age or older. Although screening among Black women increased concurrently with a greater proportion graduating from high school in the 40 to 44 and 50 to 54 age cohorts, it is noteworthy that mammography rates rose for every age cohort whether or not rates of high school graduation rose.

Breast Physical Examination

The trend in use of breast physical examination reported in the NHIS was analyzed by Makuc et al.,[18] who detected a greater increase in breast exams among Black than White women between 1973 and 1985, resulting in higher rates among Black women in 1985.

Questions on breast physical examination asked in the 1987 NHIS were slightly different in 1990. Whether women had ever heard of breast physical examination and whether they used it for screening as opposed to diagnostic purposes were asked only in 1987. In that survey, 6.9% of women reported that they had never heard of breast physical examination (13% of women had never heard of a mammogram). Of the women who reported having a breast physical examination, only 6% said that it was not for screening purposes.[11]

Since the 1990 data do not distinguish between screening and diagnostic breast physical examinations, the rate of all exams in both years is reported. Use of breast physical examination exceeded mammography in both years. Fifty percent of women more than 40 years of age reported an annual breast physical examination by a health care practitioner in 1987, and this proportion rose to 65% in 1990.

A larger percentage of Black than White or Hispanic women reported hav-

TABLE 3—Proportion of Women Reporting a Screening Mammogram in the Last 12 Months, by Age Group, Race, and Percentage Graduating from High School: 1987 and 1990

	White,[a] %				Black,[a] %				Hispanic, %			
	1987		1990		1987		1990		1987		1990	
Age, y	Mammogram	High School	Mammogram	High School	Mammogram	High School	Mammogram	High School	Mammogram	High School	Mammogram	High School
40–44	16	86	31	88	12	72	33	80	11	...	27	56
45–49	22	81	33	84	14	67	31	67	11	...	33	46
50–54	23	79	40	81	17	57	35	65	18	...	34	47
55–59	19	73	39	75	21	47	36	49	17	...	35	42
60–64	19	71	38	72	20	47	32	45	8	...	40	38
65–69	17	63	36	70	13	37	29	36	15	...	27	32
70–74	15	56	31	61	12	27	31	27	21	...	19	28
75–79[b]	15	46	28	50	8	20	31	20	3	...	39	16
80+	8	...	17	...	3	...	22	...	23	...	25	...
Total	18	...	33	...	14	...	32	...	13	...	31	...

Sources. Educational Attainment in the United States: March 1987 and 1986[33] *and Educational Attainment in the United States: March 1991 and 1990.*[34]
[a]Non-Hispanic.
[b]Educational attainment refers to women 75 years of age and older.

ing breast physical examinations in 1987 and 1990, and their rate of screening also increased most. Although rates were lower for White and Hispanic women than for Black women, all three groups showed increased use of breast physical examination.

Reasons for Not Having Mammograms

Of women more than 40 years of age who had not had a mammogram in the last 3 years, only 7.2% in 1987 and 5.7% in 1990 had ever had a mammogram. In both 1987 and 1990, the same question was asked: What is the most important reason why you have never had a mammogram/ not had a mammogram in the past few years? The 1987 survey provided 17 precoded categories (including other and don't know), and the 1990 survey provided 14 (including other, don't know, and refused).

Women's responses are grouped into six general categories in Table 4 characterizing why they did not have a mammogram in the 3 years prior to the survey: (1) lack of physician's recommendation, (2) lack of knowledge that screening mammography is for asymptomatic women, (3) fear or discomfort, (4) expense, (5) put it off, and (6) other.

The most frequently reported reason was no need or no breast problems. Forty-eight percent of respondents in 1987 and 41% in 1990 lacked knowledge of the need for breast cancer screening. Previous studies have shown that women who do not have screening mammograms are more likely to believe that the procedure is unnecessary in the absence of symptoms than women who are screened.[30] Thirty-one percent of women reported that their physician did not recommend a mammogram.

Few women reported that they did not have a mammogram because they felt fear or discomfort, and the proportion dropped, from 2.5% to 1.7%, between 1987 and 1990. Although rates remained small, the proportion of women who cited expense as a reason for not having a mammogram more than doubled from 2.7% in 1987 to 7.2% in 1990.

Discussion

Despite encouraging recent increases in screening, two thirds of women more than 40 years of age in 1990 did not have a screening mammogram in the previous year, and 38% had *never* had a mammogram. Age, cost, and physician recommendation are key reasons why women do not undergo screening. The discussion to follow explores links between these factors and strategies to increase the rate of screening mammography.

Age is most strongly associated with breast cancer in women. Data from the Surveillance, Epidemiology, and End Results Program of the NCI show that the age-adjusted incidence rate for invasive breast cancer among women 65 to 74 years of age (399 per 100 000) was twice that of women 35 to 64 years of age (193 per 100 000) during 1984 to 1987.[3] Although screening rates doubled for older women, they remained below those of younger women. Benefits of mammogra-

TABLE 4—Most Important Reasons for Not Having a Mammogram: 1987 and 1990

Group/Reason	1987, %[a]	1990, %[a]
Group 1: Lack of physician's recommendation		
Not recommended by physician	29.0	30.7
Physician said not needed	1.9	...[b]
Subtotal	30.9	30.7
Group 2: Lack of knowledge		
Didn't know I should	3.8	1.5
Not needed/not necessary	27.9	22.3
Haven't had a problem/no problem	14.2	14.3
Don't see a physician	1.9	1.8
Never heard of a mammogram	...[b]	...[b]
Subtotal	47.8	41.2
Group 3: Fear		
Fear	0.6	1.7
Fear of radiation	1.2	...[b]
Painful procedure	0.5	...[b]
Subtotal	2.3	1.7
Group 4: Expense		
Costs too much	2.5	7.2
No insurance	0.2	...[b]
Subtotal	2.7	7.2
Group 5: Put it off	6.5	9.1
Group 6: Other		
Too young	...[b]	0.3
Breasts missing	...[b]	0.6
Other	4.7	4.2
Don't know	4.6	5.0
Too embarrassing	0.1	...[b]
Unpredictable results	0.1	...[b]
Subtotal	9.5	10.1
Total	99.7	100.0

[a]Weighted frequency (SUDAAN).
[b]Question not asked that year.

phy screening for women 50 to 74 years of age have been proven in clinical trials.[31] The benefits for women 75 years of age or older are unknown; however, a 1990 forum for breast cancer screening in older women recommended annual breast physical examination and biennial mammography for women in this age group in good general health.[32] Given the finding that strong physician recommendation is the most important health education intervention for breast cancer screening,[33] especially for older women,[34] it is critical for practitioners to be aware of and convey this information to their patients and to recommend regular mammograms accordingly.

Logistic regression analyses of NHIS data from 1987[13] and 1990,[22] as well as other surveys,[35,36] support our findings that gaps by race and ethnicity closed between 1987 and 1990, leaving income and education to explain the significant differences in screening mammography in 1990. A study conducted in Los Angeles also found virtually the same rates of screening mammography for high-SES White, Black, and Hispanic women, although Hispanic women were screened at lower rates than Black or White women in the low-SES group.[12] Providers could help increase rates of mammography use, but they are often unaware of insurance coverage or state benefits for screening mammography.[37,38]

This complex picture shows subgroups of women with different rates of screening mammography, underlying beliefs about mammography, and ability to pay. The average charge for a screening mammography is between $100 and $125.[39] Yet radiologists can deliver high-quality screening mammography at much lower cost (between $50 and $65).[40–42]

Although few women reported cost as the major reason that they did not have a mammogram, it may be more important than NHIS data suggest since the survey question precludes reporting cost unless it is the foremost concern. Other studies have found that women's concerns about the cost of mammography[12] and willingness to pay $75 to $100[33] are significant in their decision to have regular mammograms. Physicians may fail to recommend mammography if, in their judgment, a woman will not comply or is not able to pay.[9] A study of breast cancer screening use by Latina community health center clients underscores the importance of providers in initiating discussion and encouraging regular screening.[43] The small proportion of women who reported cost as a deterrent in the NHIS may reflect the phenomenon that providers do not recommend screening mammograms to patients they believe will not comply.

Women are more likely to undergo regular mammography screening if they believe that their doctors support it[26,30,33,34] or think that the test is efficacious for and accessible to them.[26,30,33] Surveys that directly measured whether a physician's recommendation promoted mammography use among women found that 97%[8] and 94%[44] of respondents had a mammography after their doctor recommended it; a survey of poor urban Black women found that 85% would have a mammogram if their physician recommended one.[45] Nevertheless, most (63%) physicians queried in 1989 indicated that they did not recommend screening mammography.[2]

Most women with at least one mammogram report that they had it because their physician so recommended.[26,46] Strategies to promote regular use of mammography need to target both women and physicians.[33] Self-referral to mammography facilities is unusual; the NCI National Mammography Survey found that only 7.7% of all mammography clients were self-referred in 1991.[47]

Primary care physicians serve as gatekeepers in deciding whether to recommend procedures to their patients. Physicians may effectively filter out women with low income and education by not discussing or recommending mammography to them. Hence, in Table 4, both lack of physician recommendation and women's lack of knowledge regarding screening mammography constitute missed clinical opportunities for physicians or other health care providers.

If the procedure is to be successful, women need to understand the importance of routine screening for asymptomatic women. Training in breast physical examination is needed for physicians[48,49] as well as for the women themselves.[50,51] Physicians also need to include such examinations in their regular physicals.[16]

A recent review of the state legislative movement to make screening mammography a basic health insurance benefit argues that a clear view of attitudes toward mammography requires that physicians and age-eligible women take full advantage of benefits.[52] This requires a health care delivery system that provides confidence to primary caregivers that patient referrals will yield good care and ensures universal access to low-cost mammography screening facilities as well as careful record keeping with a reminder system.

Four components seem necessary to sustain the increases in screening mammography that occurred from 1987 to 1990: culturally specific public health programs to promote screening for early detection of breast cancer, greater acceptance of mammography as a breast cancer prevention strategy in the medical community, reorganization of facilities to provide low-cost screening mammography, and mandated universal coverage for mammogram services and breast physical examinations. □

References

1. American Cancer Society. Survey of physicians' attitudes and practices in early cancer detection. *CA.* 1985;35:197–213.
2. 1989 survey of physicians' attitudes and practices in early cancer detection. *CA.* 1990;40:77–101.
3. Coleman EA, Feuer EJ. Breast cancer screening among women from 65 to 74 years of age in 1987–88 and 1991. *Ann Intern Med.* 1992;117:961–966.
4. Use of mammography—United States, 1990. *MMWR.* 1990;39:621, 627–630.
5. Shapiro S, Strax P, Venet L. Periodic breast cancer screening in reducing mortality from breast cancer. *JAMA.* 1971;215: 1777–1785.
6. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *JNCI.* 1982;69:349–355.
7. Shapiro S, Venet W, Strax P, Venet L. *Periodic Screening for Breast Cancer; The Health Insurance Plan Project and Its Sequelae. 1963–1986.* Baltimore, Md: Johns Hopkins University Press; 1988.
8. Howard J. Using mammography for cancer control: an unrealized potential. *CA.* 1987;37:33–48.
9. Smith RA, Haynes S. Barriers to screening for breast cancer. *Cancer.* 1992;69:1968–1978.
10. Zapka JG, Hosmer D, Costanza ME, Harris DR, Stoddard A. Changes in mammography use: economic, need, and service factors. *Am J Public Health.* 1992;82:1345–1351.
11. Dawson DA, Thompson GB. Breast cancer risk factors and screening: United States, 1987. *Vital Health Stat [10].* 1989; 172. DHHS publication PHS 90-1500.
12. Stein JA, Fox SA, Murata PJ. The influence of ethnicity, socioeconomic status, and psychological barriers on use of mammography. *J Health Soc Behav.* 1991;32: 101–113.
13. Jepson C, Kessler LG, Portnoy B, Gibbs T. Black-White differences in cancer prevention knowledge and behavior. *Am J Public Health.* 1991;81:501–504.
14. Calle EE, Flanders WD, Thun MJ, Martin LM. Demographic predictors of mammography and Pap smear screening in US women. *Am J Public Health.* 1993;83:53–60.
15. Harris RP, Fletcher SW, Gonzalez JJ, et al. Mammography and age: are we targeting

the wrong women? A community survey of women and physicians. *Cancer.* 1991;67: 2010–2014.
16. NCI Breast Cancer Screening Consortium. Screening mammography: a missed clinical opportunity? *JAMA.* 1990;264:54–58.
17. Burg MA, Lane DS, Polednak AP. Age group differences in the use of breast cancer screening tests. The effects of health care utilization and socioeconomic variables. *J Aging Health.* 1990;2:514–530.
18. Makuc DM, Freid VM, Kleinman JC. National trends in the use of preventive health care by women. *Am J Public Health.* 1989; 79:21–26.
19. Polednak AP, Lane DS, Burg MA. Risk perception, family history, and use of breast cancer screening tests. *Cancer Detect Prev.* 1991;15:257–263.
20. Zapka JG, Stoddard AM, Costanza ME, Greene HL. Breast cancer screening by mammography: utilization and associated factors. *Am J Public Health.* 1989;79:1499–1502.
21. Bernstein AB, Thompson GB, Harlan LC. Differences in rates of cancer screening by usual source of medical care. Data from the 1987 National Health Interview Survey. *Med Care.* 1991;29:196–209.
22. Rakowski W, Rimer BK, Bryant SA. Integrating behavior and intention for the study of mammography: data from the 1990 supplement to the National Health Interview Survey. *Public Health Rep.* In press.
23. Massey JT, Moore TF, Parsons VL, Tadros W. Design and estimation for the National Health Interview Survey, 1989; 1985–94. *Vital Health Stat [2].* 1989;119. DHHS publication PHS 89-1384.
24. National Center for Health Statistics, Schoenborn CA, Marano M. Current estimates from the National Health Interview Survey: United States, 1987. *Vital Health Stat [10].* 1988;166. DHHS publication PHS 88-1594.
25. Adams PF, Benson V. Current estimates from the National Health Interview Survey, 1990. *Vital Health Stat [10].* 1991;181. DHHS publication PHS 92-1509.
26. Horton JA, Romans MC, Cruess DF. Mammography attitudes and usage study, 1992. *Wom Health Issues.* 1992;2:180–186.
27. Kessler L. Surveillance of breast cancer screening: trends, barriers and monitoring progress. Presented at the American Cancer Society National Conference on Breast Cancer; August 1993; Boston, Mass.
28. *SUDAAN Survey Data Analysis Software.* Research Triangle Park, NC: Research Triangle Institute; 1991.
29. Brown ML, Kessler LG, Rueter FG. Is the supply of mammography machines outstripping need and demand? An economic analysis. *Ann Intern Med.* 1990;113:547–552.
30. Rimer BK, Keintz MK, Kessler HB, Engstrom PF, Rosan JR. Why women resist screening mammography: patient-related barriers. *Radiology.* 1989;172:243–246.
31. Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. *Report of the International Workshop on Screening for Breast Cancer, February 24–25, 1993.* Bethesda, Md: National Cancer Institute; 1993:1–38.
32. Screening recommendations of the Forum Panel. *J Gerontol.* 1992;47(special issue):5.
33. Rimer BK, Trock B, Engstrom PF, Lerman C, King E. Why do some women get regular mammograms? *Am J Prev Med.* 1991;7:69–74.
34. Coll PP, O'Connor PJ, Crabtree BF, Besdine RW. Effects of age, education, and physician advice on utilization of screening mammography. *J Am Geriatr Soc.* 1989; 37:957–962.
35. Farley TA, Flannery JT. Late-stage diagnosis of breast cancer in women of lower socioeconomic status: public health implications. *Am J Public Health.* 1989;79: 1508–1512.
36. Lane DS, Polednak AP, Burg MA. Breast cancer screening practices among users of county-funded health centers vs women in the entire community. *Am J Public Health.* 1992;82:199–203.
37. Mayer JA, Slymen DJ, Jones JA, et al. Mammography knowledge and intentions among insured women. *Prev Med.* 1992; 21:8–17.
38. Boss LP, Guckes FH. Medicaid coverage of screening tests for breast and cervical cancer. *Am J Public Health.* 1992;82:252–253.
39. Physician Payment Review Commission. *The Cost of Providing Mammography. Report to Congress.* Washington, DC: US Government Printing Office; 1989.
40. Bassett LW, Fox SA, Pennington E, Gold RH. Screening mammography in southern California: 2-1/2 year longitudinal survey of fees. *Radiology.* 1989;173:61–63.
41. Sickles EA, Weber WN, Galvin HB, Ominsky SH, Sollitto RA. Mammographic screening: how to operate successfully at low cost. *Radiology.* 1986;160:95–97.
42. Bird RE, McLelland R. How to initiate and operate a low-cost screening mammography center. *Radiology.* 1986;161:43–47.
43. Zapka JG, Stoddard A, Costanza ME, Mas E. Breast cancer screening utilization by Latina community health center clients. *Health Educ Res.* 1989;4:461–468.
44. *A Study of Women's Awareness and Use of Mammograms.* New York, NY: Lieberman Research Inc; 1987.
45. Price JH, Desmond SM, Slenker S, Smith D, Stewart PW. Urban Black women's perceptions of breast cancer and mammography. *J Community Health.* 1992;17:191–204.
46. Fox SA, Murata PJ, Stein JA. The impact of physician compliance on screening mammography for older women. *Arch Intern Med.* 1991;151:50–56.
47. Houn F, Brown ML. The current practice of screening mammography in the U.S.: data from the National Survey of Mammography Facilities. *Radiology.* In press.
48. Fletcher SW, O'Malley MS, Bunce LA. Physicians' abilities to detect lumps in silicone breast models. *JAMA.* 1985;253: 2224–2228.
49. Fletcher SW, Fletcher RH. The breast is close to the heart. *Ann Intern Med.* 1992; 117:969–971. Editorial.
50. Coleman EA. Practice and effectiveness of breast self examination: a selective review of the literature (1977–1989). *J Cancer Educ.* 1991;6:83–92.
51. Coleman EA, Riley MB, Fields F, Prior B. Efficacy of breast self-examination teaching methods among older women. *Oncol Nurs Forum.* 1991;18:561–566.
52. McKinney MM, Marconi KM. Legislative interventions to increase access to screening mammography. *J Community Health.* 1992;17:333–349.
53. *Educational Attainment in the United States: March 1987 and 1986.* Table 2: "Years of school completed by persons 15 yrs old & over, by single years of age, sex, race, and Hispanic origin: March 1987." Washington, DC: US Bureau of the Census; 1988. US Bureau of the Census Current Population Report, Series P-20, No. 428.
54. *Educational Attainment in the United States: March 1991 and 1990.* Table 1: "Years of school completed by persons 15 yrs old & over, by age, sex, race, and Hispanic origin: March 1991 and 1990." Washington, DC: US Bureau of the Census; 1992. US Bureau of the Census Current Population Report, Series P-20, No. 462.

The Next Generation: Poor Compliance with Risk Factor Guidelines in the Children of Parents with Premature Coronary Heart Disease

ABSTRACT

Objectives. The offspring of individuals with premature coronary heart disease are themselves at increased risk for myocardial infarction before the age of 55. Consensus panels have recommended that all such offspring undergo an evaluation of cardiovascular risk, including cholesterol testing.

Methods. To examine self-reported rates of cardiovascular risk factor assessment in this population, we conducted a telephone survey of 318 Canadian adults with premature coronary heart disease and of one offspring from 298 (94%) of the 318 families.

Results. The median age of the offspring was 20 years (range 2 to 39 y). Among the 219 late adolescent and young adult offspring, only 97 (44%) reported having had a blood cholesterol measurement during the preceding 3 years. Thirty-seven percent reported being current smokers, 31% were overweight, and 30% exercised fewer than three times per week. Men were less likely than women to report having had their blood pressure measured in the preceding year (57% vs 80%).

Conclusions. These low rates of cardiac risk factor assessment in families of patients with premature coronary heart disease represent missed opportunities for primary prevention. More effective strategies to prevent atherosclerosis in this population are needed. (*Am J Public Health.* 1994;84:68–71)

Neima R. Langner, MD, Peter C. Rowe, MD, and Richard Davies, MD

Introduction

Individuals with a history of premature coronary heart disease in one or more first-degree relatives are 2 to 13 times more likely to develop a myocardial infarction before the age of 55 than are members of the general population.[1,2] Moreover, risk factors for coronary heart disease aggregate within families,[3–7] and many offspring of individuals with premature atherosclerotic heart disease have unfavorable lipoprotein profiles.[8–11] Based on these observations, the Canadian Pediatric Society, the American Academy of Pediatrics, and consensus panels in both Canada and the United States have recommended that all offspring of parents with premature coronary heart disease undergo an evaluation of cardiovascular risk, including cholesterol or lipoprotein testing.[12–16] To assess whether these recommendations are being adopted, we surveyed Ottawa residents with premature coronary heart disease and a sample of their offspring regarding their awareness and management of cardiovascular risk factors.

Methods

Definitions

Coronary heart disease was defined as a documented myocardial infarction or at least a 50% diameter stenosis on coronary angiography, as graded visually by the angiographer. Criteria for the diagnosis of myocardial infarction were those used in the Cardiac Arrhythmia Suppression Trial.[17,18] Coronary heart disease was classified as premature if it occurred in men younger than 50 years of age or in women younger than 60.

Study Population

Patients meeting the age criteria were eligible for this study if they lived in the Ottawa metropolitan area and had either suffered a myocardial infarction or undergone coronary angiography between January 1, 1988, and February 28, 1990. We identified eligible patients from (1) the computerized records of the University of Ottawa Heart Institute catheterization laboratory, the only cardiac catheterization laboratory for adults in the city; and (2) a computerized database of patients admitted to seven regional hospitals with a new myocardial infarction (eligible participants for the Cardiac Arrhythmia Suppression Trial).

Sixty individuals had participated in the Cardiac Arrhythmia Suppression Trial and were exposed to more intensive medical surveillance than the general population of those with premature coronary heart disease. Because this could have increased the rate of risk factor assessment among both patients and their offspring, trial participants were excluded from this survey. The remaining patients screened for participation but not enrolled in the trial included 408 age-eligible patients suffering a myocardial infarction between January 1, 1988, and February 28, 1990. Of these, 73 lived outside the Ottawa region, 7 had clinically significant preexisting non-ischemic heart disease, and 77 lacked a recorded telephone number. We

At time of the study, the authors were with the Departments of Cardiology and of Epidemiology and Community Medicine and with the Ottawa Heart Institute at the University of Ottawa School of Medicine, Ottawa, Canada. Dr Rowe is now with the Department of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore, Md.

Requests for reprints should be sent to Peter C. Rowe, MD, Department of Pediatrics, Brady 212, Johns Hopkins Hospital, 600 N Wolfe St, Baltimore, MD 21287.

This paper was accepted June 15, 1993.

attempted to contact the remaining 251 patients.

Of the 5643 diagnostic angiograms performed in the same 2-year period, 1141 patients met both age and angiographic criteria for this study. We excluded 35 who lacked a telephone and 115 who had already been identified from the Cardiac Arrhythmia Suppression Trial database. From the remaining 991, we selected a random sample of 356 patients using a computer-generated list of random numbers.

Design

Interviewers obtained verbal consent before proceeding with the telephone interview. After each adult with premature coronary heart disease was interviewed, we sought permission to contact one offspring for a similar interview. Eligible offspring were those over 2 years of age. From those eligible, we selected the individual in each family whose birthday was closest to and before the date of the parent interview. We interviewed offspring aged 16 years and older directly, and asked the parent most familiar with the child's habits to respond for children under 16. The study was approved by the Research Ethics Committee of the Ottawa Civic Hospital.

Study Instrument

The questionnaire for this study was developed using both original questions and questions previously validated in larger health surveys. The structured telephone interview consisted primarily of standardized short-answer questions on the following topics: (1) general health, (2) family history of heart disease (elicited from the parent), (3) dietary changes made in the previous year, and (4) self-reported cholesterol and blood pressure measurements.

The questionnaire was pretested for clarity, content coverage, and face validity. Eight interviewers (medical students and medical secretaries) conducted the surveys over a 12-week period beginning in July 1990. Six telephone calls at different times of the day were attempted over a 2-week period before eligible individuals were regarded as unreachable. If the index patient had died, no attempt was made to interview the offspring. We did not attempt to locate subjects whose telephone numbers were incorrectly listed on the computer database.

Sample Size Calculations

We estimated that 20% of the target group would be ineligible (lack of offspring or dead), that 20% would be unreachable, and that 10% would refuse to be interviewed. The chief outcome measure in this study was the proportion of offspring who had reported having undergone blood cholesterol testing. We estimated this proportion to be 25%. To obtain a 95% confidence interval within 5% of this figure, we determined that we would need to approach approximately 600 patients from both databases and to interview 300, using a sampling frame of one offspring per patient.

TABLE 1—Response Rates, Exclusions, and Reasons for Nonresponse in the Total Sample

	No.
Selected sample	607
Ineligible	
No children	57
Deceased[a]	16
Age incorrect	2
Parent interview only	21
Parent and child interview completed	297
Nonrespondents	
Wrong phone number[b]	120
No answer	29
Refused	48
Other[c]	17

[a]One offspring of a deceased patient was interviewed.
[b]"Wrong phone number" includes those who had moved, whose numbers were not in service, or whose numbers had been entered incorrectly into the database.
[c]"Other" includes respondent's inability to speak French or English, or parent or child being in the hospital or out of the country.

TABLE 2—Risk Factors among Parents with Premature Coronary Heart Disease (n = 318)

	No.	%
Male	204	64.2
Smokers[a]	99	31.1
Ex-smokers (all)	174	54.7
Quit < 5 years (% of ex-smokers)	125	71.8
Overweight, by BMI[b]		
25–26.9	63	19.8
≥27	172	54.1
Hypertension	116	36.5
High cholesterol	175	55.0
Family history[c]		
One relative	112	35.2
Two or more relatives	64	20.1

[a]Defined as those who had smoked at least 100 cigarettes and were still smoking.
[b]BMI = body mass index (weight/height^2).
[c]Positive family history denotes relative with heart attack before age 55.

Data Analysis

Comparisons between groups were conducted using contingency tables and chi-square statistics for categorical variables, and two-tailed *t* tests or analysis of variance for interval or continuous data. The Fisher's Exact Test was used for categorical analysis when warranted by small cell sizes.

Results

Description of the Study Population

Table 1 shows the response rates, exclusions, and reasons for nonresponse. The refusal rate among contacted eligible patients was 13% (48/367). Interviewed subjects did not differ from those who were not interviewed or not surveyed in either the prevalence of cardiovascular risk factors (smoking, diabetes, previous myocardial infarction) or angiographic measures of heart disease severity (left ventricular dysfunction, left main disease, triple vessel disease).

Table 2 shows the distribution of risk factors for coronary heart disease in this population. Fifty-two percent had three or more of these risk factors. We interviewed one offspring from 298 (94%) of the 318 families. Among the offspring, 56% were female; 10% were aged 2 to 10 years, 36% were age 10 to 19, 40% were age 20 to 29, and 14% were age 30 to 39.

Risk Factors in Offspring Older Than 16 Years

In addition to the positive family history for premature coronary heart disease, information was available on five other classic risk factors from the 219 offspring aged 16 years and older (Table 3). By these criteria, 165 (75%) had at least one additional risk factor, 88 (40%) had only one, 63 (29%) had two, and 14 (6%) had three or more. Table 4 displays comparable results for both the general population of Ottawa in 1987 and those with a family history of heart disease.

Of respondents aged 16 years and older, 78% had been examined by a physician in the preceding 3 years. Only 97 (44%) reported having had a blood cholesterol measurement performed during those 3 years, and just 5 of these 97 actu-

TABLE 3—Reported Classic Risk Factors for Ischemic Heart Disease among Progeny 16 Years of Age or Older

	Male (n = 93)	Female (n = 126)	Total (n = 219)		P
	%	%	%	95% Confidence Interval	(Male vs Female)
Smokers[a]	45	32	37	31, 44	.06
Ex-smokers	9	17	13	9, 18	.12
Overweight, by BMI[b]	41	24	31	25, 37	<.01
25–26.9	24	6	13	9, 18	<.001
≥27	17	18	18	13, 23	.98
Hypertension	5	14	11	7, 15	.11
High cholesterol	7	11	9	6, 14	.34
Exercise < three times/week	16	40	30	24, 36	<.001

[a]Defined as those who had smoked at least 100 cigarettes and were still smoking.
[b]BMI = body mass index (weight/height^2).

TABLE 4—Risk Factors, Measurement, and Changes: Comparison of Current Survey and Ottawa Carleton Heart Beat Survey (1987)[a]

	Current Survey, %	Heart Beat Survey, FH CHD[b]		P[c]
		Yes, %	No, %	
Feel health is fair or poor	8.9	6.1	5.0	.197
Had blood pressure checked within 1 year	74.5	78.8	78.5	.488
Had cholesterol checked within 1 year	38.2	22.7	25.3	.001
Was informed of high blood pressure	13.3	16.6	13.6	.490
Get enough exercise	36.7	26.5	33.3	.051
Exercise > 15 minutes three times/week	68.3	54.8	60.3	.016
Smoking history				
Never smoked	44.4	47.7	48.0	.714
Quit smoking	15.6	20.1	14.5	.146
Current smoker	40.0	32.3	37.5	.193
Tried to stop	68.1	71.3	66.5	.729
Body mass index ≥ 25	35.2	29.6	24.9	.038
Trying to lose weight				
Total	49.7	46.0	40.1	.070
Men	33.3	35.7	27.7	.209
Women	58.4	54.6	50.0	.368
Body mass index ≥ 25	61.9	67.1	57.6	.424

Source. Data accessed with permission of Dr G. Dunkley.
[a]Respondents were 19 to 39 years of age. Thus, for the Heart Beat Survey, n = 679; for the current survey, n = 178.
[b]FH CHD = Family history of coronary heart disease (yes = 279, no = 400), as defined by the response to the question: "Has anyone in your immediate family had heart problems?"
[c]Correspond to chi-square values, which represent a 3 × 2 comparison based on the three data columns.

ally knew their cholesterol levels. Offspring were more likely to have had a cholesterol measurement performed if they were at least 25 years of age (60% vs 42%; $P < .01$) and if they had undergone a physical examination within the past year (59% vs 36%; $P = .001$).

Knowledge of Risk Factors

When asked to list the major causes of heart disease or heart attacks, respondents most frequently cited dietary factors (mentioned by 99%), smoking (58%), stress (57%), lack of exercise (21%), and high blood pressure (9%). Only 13% identified heredity as an important factor. Current smokers were more likely to rate smoking as a cause of heart attacks than were nonsmokers (72% vs 59%; $P < .01$).

Changes in Risk Factors

Among offspring at least 16 years of age, 67% (60% of men and 72% of women) reported having changed their diet in the preceding 2 years to reduce their risk of heart disease. Univariate analyses showed no clinically or statistically significant difference between those who did and did not change their diets with regard to age, education, marital status, smoking status, body mass index, physical exam within the previous year, or blood pressure measurement within the previous year. Only the level of worry about future heart disease was significantly related to dietary change. Of those who had a below-average worry rating, only 52% had changed their diets, compared with 76% of those who worried more ($\chi^2 = 13.5$, $P = .001$). Smokers were almost three times as likely as nonsmokers to report having made no efforts to improve their health in the preceding year (30% vs 11%, $P = .001$). Those with two risk factors for coronary heart disease besides family history were marginally less likely to have made changes in diet in the preceding year than were those who had no other risk factors (77% vs 91%, $P < .04$).

Offspring Younger Than Age 16 Years

More limited information was available regarding the remaining 79 offspring aged 2 to 16 years. Nine parents provided proxy responses for their 16-year-old children; we analyzed data for these adolescents with the younger offspring. Fifty-five children (70%) had seen a physician in the preceding year. Sixteen parents (20%) reported that the child had undergone a cholesterol measurement at some time. Forty-four parents (56%) were aware of a measurement of the child's blood pressure in the preceding year.

Discussion

This survey of the offspring of Canadian adults with premature coronary heart disease has identified a low rate of compliance with current guidelines for the prevention of atherosclerosis.[12,13] Only 38% (113/298) of the offspring reported being aware that they had undergone cholesterol testing in the preceding 3 years, and only 57% of those aged 16 and over reported having had their blood pressure measured in the preceding year. Many of the late adolescent and young adult offspring also reported being smokers, being overweight, and exercising less often than three times per week. That 37% were current smokers is of particular concern in light of evidence that smoking multiplies the risk of ischemic heart disease in those with a positive family history.[19,20]

These data show that the current case-finding approach to identifying and screening high-risk offspring is ineffective. Although the reasons for this are complex, we identified a lack of awareness of the increased risk posed by a positive family history, as has been reported in studies of siblings of adults with coronary heart disease.[5] While it is true that the offspring in our survey were not asked to rate the relative importance of family history as a contributor to the development of heart disease, only 13% of the older offspring mentioned heredity as a major cause. Moreover, only 55% of the parents with premature coronary heart disease reported premature myocardial infarction among their close relatives. It thus appears that this increased risk has received insufficient attention over two generations, and that the attitudes and behaviors of the parents regarding cardiovascular risk are shared by their children.

A comparison of the present survey with the 1987 Ottawa Carleton Heart Beat Survey suggests that recommendations for the primary prevention of premature coronary heart disease are being adopted at a slow pace, as has been observed in the United States.[21] Compared with findings from the Heart Beat Survey, we found substantially higher rates of cholesterol screening. Although this difference may reflect more intensive screening of family members than of others in the population, as was recommended in Canada in 1989,[22] it is also possible that more people in 1990 were having cholesterol screening regardless of family history.

We cannot attribute the low rates of cardiovascular risk assessment identified in this study to a lack of public health information about the link between elevated cholesterol levels and heart disease. During the decade before this survey, residents of Ottawa were exposed to local heart health campaigns as well as to intense national and US media coverage of this issue. Unlike the United States at present, there are no financial barriers to medical examination and lipoprotein measurement in the Canadian health care system. However, there may be disincentives in both countries to identifying and providing preventive care to those with a history of premature coronary heart disease among first-degree relatives. To function well, the present system demands one of two things: either an exemplary effort by practitioners to schedule assessments of offspring (some of whom may not be members of the physician's practice) once premature coronary heart disease is recognized in a parent; or an unusually high level of knowledge among offspring about the risk associated with their family history, which then leads them to request an appointment. An alternative to the current method would be a structured program of education, risk factor assessment, and intervention, which would be offered to family members of patients presenting with premature coronary heart disease. Before such a program could be implemented, however, it would have to be demonstrated that the preventive strategy leads to the desired changes in behavior without adding an unhealthy psychological burden to the offspring identified as being at increased risk. □

Acknowledgments

This study was supported by a grant from Merck-Frosst Canada, Inc. Dr Rowe was supported by a Career Scientist Award from the Ministry of Health of Ontario.

We thank Drs Nicholas Birkett, Catherine Neill, Barton Childs, and Diane Becker for their comments on an earlier version of this manuscript. We are grateful to Dr Geoffrey Dunkley for permission to use data from the Ottawa Carleton Heart Beat Survey.

References

1. Nora JJ, Lortscher RH, Spangler RD, Nora AH, Kimberling WJ. Genetic-epidemiologic study of early-onset ischemic heart disease. *Circulation.* 1980;61:503–508.
2. Hunt SR, Williams RR, Barlow GK. A comparison of positive family history definitions for defining risk of future disease. *J Chronic Dis.* 1986;39:809–821.
3. Slack J, Evans KA. The increased risk of death from ischaemic heart disease in first degree relatives of 121 men and 96 women with ischaemic heart disease. *J Med Genet.* 1966;3:329–357.
4. Phillips RL, Lilienfeld AM, Diamond EL, Kagan A. Frequency of coronary heart disease and cerebrovascular accidents in parents and sons of coronary heart disease index cases and controls. *Am J Epidemiol.* 1974;100:87–100.
5. Becker DM, Becker LC, Pearson TA, Fintel DJ, Levine DM, Kwiterovich PO. Risk factors in siblings of people with premature coronary heart disease. *J Am Coll Cardiol.* 1988;12:1273–1280.
6. ten Kate LP, Bowman H, Daiger SP, Motulsky AG. Familial aggregation of coronary heart disease and its relation to known genetic risk factors. *Am J Cardiol.* 1982; 50:945–953.
7. Perkins KA. Family history of coronary heart disease: is it an independent risk factor? *Am J Epidemiol.* 1986;124:182–194.
8. Glueck CJ, Fallat RW, Tsang R, Buncher CR. Hyperlipemia in progeny of parents with myocardial infarction before age 50. *Am J Dis Child.* 1974;127:70–75.
9. Lee J, Lauer RM, Clarke WR. Lipoproteins in the progeny of young men with coronary artery disease: children with increased risk. *Pediatrics.* 1986;78:330–337.
10. Sveger T, Fex G, Borgfors N. Hyperlipidemia in school children with family histories of premature coronary heart disease. *Acta Paediatr Scand.* 1987;76:311–315.
11. Dennison BA, Kikuchi DA, Srinivasan SR, Webber LS, Berenson GS. Parental history of cardiovascular disease as an indication for screening for lipoprotein abnormalities in children. *J Pediatr.* 1989;115: 186–194.
12. Children's diets and atherosclerosis. *Can Med Assoc J.* 1981;124:1545–1548.
13. Prudent life-style for children: dietary fat and cholesterol. *Pediatrics.* 1986;78:521–524.
14. Committee on Nutrition, American Academy of Pediatrics. Indications for cholesterol testing in children. *Pediatrics.* 1989; 83:141–142.
15. Canadian consensus conference on cholesterol: final report. *Can Med Assoc J.* 1988; 139(11, suppl):1–8.
16. Lowering blood cholesterol to prevent heart disease. *JAMA.* 1985;253:2080–2086.
17. Cardiac Arrhythmia Suppression Trial Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med.* 1989;321:406–412.
18. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: the Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* 1991;324:781–788.
19. Hopkins PN, Williams RR, Hunt SC. Magnified risks from cigarette smoking for coronary prone families in Utah. *West J Med.* 1984;141:196–202.
20. Khaw K, Barrett-Conner E. Family history of heart attack: a modifiable risk factor? *Circulation.* 1986;74:239–244.
21. Frank E, Winkleby MA, Fortmann SP, Rockhill B, Farquhar JW. Improved cholesterol-related knowledge and behavior and plasma cholesterol levels in adults during the 1980s. *JAMA.* 1992;268:1566–1572.
22. Toronto Working Group on Cholesterol Policy. *Detection and Management of Asymptomatic Hypercholesterolemia.* Policy document prepared for the Task Force on the Use and Provision of Medical Services, Ontario Ministry of Health and Ontario Medical Association, Toronto, Canada; 1989.

"Eat For Health": A Supermarket Intervention for Nutrition and Cancer Risk Reduction

ABSTRACT

Objectives. The results of an evaluation of "Eat for Health," a supermarket nutrition intervention, are presented. The study tested whether such a program could be successfully carried out and whether it could effect changes in knowledge, attitudes, and food purchasing behavior in line with nutrition and cancer risk reduction guidelines.

Methods. The evaluation consisted of an in-store monitoring element, an in-store and telephone consumer survey, and an analysis of sales data on selected foods. A matched-pair design, using a total of 40 stores in the intervention and comparison groups, was used.

Results. The intervention was successfully implemented and had limited success in changing some food purchasing behaviors. There appeared to be no effect on knowledge and attitudes except for increased awareness of a link between diet and cancer and of the program itself.

Conclusions. Despite the intervention's success, limitations of the consumer survey and sales data analyses and the continuing diffusion of nutrition messages throughout society make it difficult to specify the impact of this program on consumer nutrition knowledge and behaviors. (*Am J Public Health.* 1994;84:72–76)

Anne Brown Rodgers, Larry G. Kessler, ScD, Barry Portnoy, PhD, Arnold L. Potosky, MHS, Blossom Patterson, MA, Janet Tenney, RD, MS, Frances E. Thompson, MPH, PhD, Susan M. Krebs-Smith, RD, MPH, PhD, Nancy Breen, PhD, Odonna Mathews, MBA, and Lisa L. Kahle

Introduction

Since the 1970s, a variety of programs have been implemented to test the proposition that information provided in supermarkets not only can increase awareness and knowledge of healthful food choices but can provide specific, focused guidance in making those choices. Evaluations of the few programs with evaluation components have shown mixed results, with some improvement in knowledge and attitudes and modest, if any, changes in behavior.[1–13]

In March 1987, the National Cancer Institute (NCI), one of the National Institutes of Health, launched a 2-year multicomponent supermarket intervention in cooperation with Giant Food Inc, a major supermarket chain in the Washington, DC–Baltimore, Md, area. The "Eat for Health" program was intended to stimulate changes in knowledge and food purchasing behavior consistent with NCI's dietary recommendations for cancer risk reduction. These recommendations focus on reducing fat intake and increasing consumption of fruit, vegetables, and other fiber-containing foods.[14] The intervention was conducted in Giant's 105 Washington-area stores, while its 30 Baltimore stores served as the comparison group. Program elements included special shelf price labels; a food guide containing calorie, fat, cholesterol, sodium, and fiber values for all items carrying the special shelf price labels; a monthly bulletin containing nutrition information and recipes; signs in the produce department; and an intensive multimedia advertising campaign. Individual 3-month interventions, which were evaluated separately, were developed during the second year of the campaign in conjunction with the National Livestock and Meat Board and several poultry producers. The components and development of Eat for Health are described more fully in Light et al.[15]

The Eat for Health study was evaluated to (1) assess participating stores' ability to implement this multimethod nutrition education program; (2) measure changes in consumer purchases of fiber-rich and lower-fat foods; and (3) assess changes in consumer knowledge about diet and health issues, specifically diet and cancer issues, and self-reported food purchasing and preparation.

Methods

The Eat for Health evaluation consisted of an in-store monitoring program, an in-store and telephone consumer survey, and an analysis of sales data on selected foods.

A matched-pair design was selected, using 20 stores in the intervention group and 20 stores in the comparison group. Data from the 1980 US Bureau of the Cen-

Anne Brown Rodgers is a consultant to Prospect Associates, Rockville, Md. Larry G. Kessler, Barry Portnoy, Arnold L. Potosky, Blossom Patterson, Frances E. Thompson, Susan M. Krebs-Smith, and Nancy Breen are with the Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Md. Janet Tenney and Odonna Mathews are with the Consumer Affairs Department, Giant Food Inc, Washington, DC. Lisa L. Kahle is with Information Management Services Inc, Silver Spring, Md.

Requests for reprints should be sent to Anne Brown Rodgers, 1525 N Kenilworth St, Arlington, VA 22205.

This paper was accepted May 11, 1993.

Editor's Note. See related editorial by Zapka (p 12) in this issue.

sus decennial survey were used to match stores with regard to their neighborhoods' racial composition, age distribution, occupation, income, and proportion of homeowners. Stores were also matched with regard to volume and presence or absence of a pharmacy.

In-Store Monitoring

The Eat for Health monitoring component consisted of nine visits to the 20 intervention stores selected for evaluation. In these visits an NCI staff person talked with the store manager and checked on availability of program materials. This component was based on experience from previous supermarket interventions that showed the importance of the availability of program materials and a commitment to the program by supermarket managers and staff.[7,9] Odenkirchen et al. have described the monitoring program in more detail.[16]

Consumer Surveys

This series of three surveys was administered to shoppers from the 20 matched pairs of Washington and Baltimore stores selected for evaluation. Wave 1 was a baseline survey conducted in February and March 1987, wave 2 was conducted in January, February, and March 1988, and wave 3 was conducted in February, March, and April 1989. The shortened version of the food frequency portion of NCI's Health Habits and History Questionnaire was adapted for mail and self-reported administration.[17]

Potential survey participants were included only if they did more than 50% of the household shopping and shopped at Giant more than 50% of the time; ate less than 50% of their meals outside the home; were between the ages of 21 and 75 years; had no medical condition requiring a special diet; and did not work in the food, health, or market research industries. Table 1 provides further detail on response rates.

To evaluate responses to the three cross-sectional surveys, an analysis of covariance approach was adopted in which the survey wave (1, 2, or 3) and place (Washington vs Baltimore) served as main effects and the survey response items served as the dependent variables in a series of ordinary least squares multiple regression models.

Sales Data

Summary sales data were obtained each week from Giant's computerized cash registers, which record purchases by Universal Product Code. Data were collected for 1 baseline year (March 1986 through February 1987) and the 2 intervention years (March 1987 through March 1989) in the same 20 matched pairs of stores used for the monitoring and consumer surveys. The analysis of sales data focused on changes in purchases of fiber-containing foods, although sales data for meat and poultry were analyzed as well. A cross-sectional time series regression procedure was used to analyze the data.

Eight mutually exclusive food categories were targeted for analysis: dry cereal, baked goods, fresh produce, frozen vegetables, canned vegetables, canned and frozen beans, dried beans, and dried fruit.

Foods in each of the eight categories tracked for analysis were grouped into two subcategories: "recommended" and "other than recommended." Recommended foods contained at least 2 g of fiber and less than 30% of calories from fat. All other foods in each category were other than recommended.

The effect of the Eat for Health intervention was measured by two variables. The first was "recommended ounces," or sales in ounces of recommended food items in each category. The second was "percent recommended," or the percentage of all the foods sold in each category that were in the recommended subcategory. This portion of the evaluation is described in Patterson et al.[18]

TABLE 1—Number of People Responding and Response Rate at Each State of Data Collection, by Wave

Response Category	Wave 1	Wave 2	Wave 3
A. Total approached	3084	3268	3218
B. Initially refused (would not speak to interviewer)	729	641	973
C. Refused (qualified but would not participate)	170	166	112
D. Disqualified	682	1141	734
E. Qualified and participated [A – (B+C+D)]	1503	1320	1399
F. Response rate, % (E/A – D)	62.6	62.1	56.3
G. Completed questionnaire	958	1045	1031
H. Response rate, % (G/E)	63.7	79.2	73.7
I. Completed telephone interview	917	1019	1042
J. Completed questionnaire and telephone interview	917	957	893
K. Response rate, % (J/G)	95.7	91.6	86.6

Results

In-Store Monitoring

At the start of the Eat for Health program, it was determined that project elements (shelf labels, food guides, produce signs, and monthly bulletins) complied with the research protocol if they were available and located properly and if the nutrition information on the shelf labels was correct. The goal for project elements was an average compliance rate of 80%.

Not surprisingly, the shelf labels had the highest compliance rate, averaging 92% to 100% over the program. This high compliance rate was largely due to the fact that the labels were computer-generated and therefore uniform and up-to-date in content. Also, once attached to the shelf the labels did not move.

The *Eat for Health Food Guide* also had a high compliance rate, with booklets available at an average of 85.3% of all checkouts. Like the shelf label, this element was stable; once placed in a rack at the checkout counter, the booklets were not moved and only had to be replaced as copies were sold.

The produce signs also met the project goal, with an average compliance rate of 83.2%. This average was considered particularly good, as the signs were frequently moved and changed when items in the produce department were shifted.

The Eat for Health monthly bulletin had the lowest compliance rate over the monitoring periods, averaging 63.5% for operating checkouts. Compliance differed most significantly from store to store, with a range of 31.6% to 89.5% at operating checkouts. This lower compliance rate was of some concern because the free take-home bulletin was the only element of the program that contained detailed information on Eat for Health issues. One possible reason may have been the bulletin's popularity. It was often out of stock by midmonth. Another reason may have been its location at the end of the checkout, a busy place that is crowded with materials such as grocery bags and other store flyers. Despite this disadvantage, previous programs had demonstrated that

TABLE 2—Results for the Five Hypotheses Developed from the Overall Objectives of the Eat for Health Consumer Surveys

General Message/Questions	Survey Items	Adjusted Percentage Difference[a] Consumption	Purchase	Comment
	Hypothesis 1			
Choose more high-fiber foods	Vegetables	1.2	3.8	No statistically significant effect except for beans; insufficient to confirm hypothesis.
	Fruit	0.6	−2.0	
	Whole grains	−1.5	0.9	
	Beans	0.5	8.3*	
	Potatoes	−0.5	5.6	
Choose less high-fat foods	Bakery/snacks	−1.1	−8.9*	Trend favors positive program effect, but insufficient to confirm hypothesis.
	Whole milk prod.	−0.9	−4.4	
	High-fat foods	−4.2*	−1.4	
	Fried food	−0.8	...	
	Fats/oils	−1.1	...	
Choose more low-fat foods	Low-fat/skim milk	−1.1	3.1	Trend favors positive program effect, but insufficient to confirm hypothesis.
	Poultry	0.4	7.6	
	Fish	−0.1	−2.2	
	Hypothesis 2			
Use more low-fat cooking methods	Baking	0.0		Insufficient evidence to confirm hypothesis.
	Broiling	8.6		
	Boiling	−2.2		
	Steaming	3.6		
	Microwaving	−3.5		
	Pressure-cooking	2.7		
Fry less; eat less charred/barbecued meats	Frying	−4.7		
	Stir-frying	−9.2*		
	Sautéing	−4.9		
	Grilling/barbecuing	−3.0		
	Hypothesis 3			
Food items that reduce cancer risk?	Fiber	3.8		Overall trend favors positive effect, although insufficient to confirm hypothesis.
	Fresh fruit	15.7*		
	Vegetables	2.2		
	Whole grains	−2.8		
	Fish/poultry	3.8		
Food items that increase cancer risk?	Dairy products	0.7		
	Fatty foods	−6.3		
	Fatty meats	8.9		
	Hypothesis 4			
Is diet related to cancer?	Unprobed	14.2*		Hypothesis well supported by results.
	Direct questions	5.4		
Is diet related to other chronic diseases?	Heart disease	−3.0		
	Obesity	6.1*		
	Diabetes	3.1		
	Hypothesis 5			
Heard of Eat for Health?	Via mass media	18.8*		Hypothesis well supported by results.
	In store	3.2		
Heard of special promotion?	Via mass media	0.1		
	In store	10.4*		
Heard of fiber?	Via mass media	8.4*		
	In store	17.0*		
Store provides nutrition information?		19.3*		
Shelf labels provide nutrition information?		11.4*		

[a]Difference between Washington and Baltimore in change from wave 1 to wave 3, controlling for 12 sociodemographic and health-related variables.
*$P < .05$.

the checkout was the best possible location for the bulletins.

Consumer Surveys

It was hoped that the five research hypotheses that were developed from the surveys would provide substantial insight into the dynamics of nutrition knowledge and practices in the Washington and Baltimore communities and would demonstrate an effect of the intervention materials on food purchasing and consumption behavior. These hopes were not entirely realized, as only two of the hypotheses were supported by the data. Table 2 summarizes the results for the five hypotheses.

Hypothesis 1. Self-reported purchasing and consumption of foods high in fiber and low in fat would increase, and purchasing and consumption of foods high in fat would decrease.

The analysis showed no statistically significant effect on self-reported purchases of high-fiber foods, with the exception of dried beans. In contrast, the direction of differences from wave 1 to wave 3 appeared to show a program effect for choosing fewer high-fat foods. However, the majority of differences were small and the results were not sufficient to confirm the hypothesis.

Hypothesis 2. There would be changes in food preparation methods consistent with the Eat for Health recommendations.

The only consistent trends that appeared here were for less frying and use of barbecued food, but these trends were not statistically significant.

Hypothesis 3. There would be an increase in knowledge about foods related to cancer risk reduction.

Although an overall pattern favoring a positive effect was apparent, it was not strong enough to confirm the hypothesis.

Hypothesis 4. There would be an increased awareness of the relationship between diet and chronic disease.

It appears that Eat for Health's diet and cancer message made a clear impression on Washington shoppers. The proportion of these shoppers who mentioned, unprompted, a link between diet and cancer increased by 14.2 percentage points over the Baltimore figure from wave 1 to wave 3. This was a large and significant change.

Hypothesis 5. Awareness of the Eat for Health program would increase.

This hypothesis was also well supported by the results. Significantly more Washington shoppers than Baltimore shoppers heard of Eat for Health over

time. In particular, awareness of fiber information and awareness of the supermarket as a source of nutrition information favored a positive program effect.

Sales Data

Dry cereals. The dry cereal category included all dry cereals sold by Giant. The recommended subcategory included all high-fiber cereals. Sales data showed a striking decline in sales of these cereals. After price and other confounding variables were taken into account, the data showed a greater decrease in sales of recommended cereals in Washington than in Baltimore, indicating a negative intervention effect.

Baked goods. The baked goods category included all types of breads and rolls. Recommended types included whole wheat, natural grain, and bran-type breads. In both intervention years, the number of recommended ounces was greater than in the baseline year, but intervention effects were negative. This indicated that shoppers in both areas bought more of the recommended breads, but the trend was stronger in Baltimore than in Washington.

Fresh produce. The fresh produce category included a wide variety of fresh fruits and vegetables available in Giant supermarkets. The number of recommended ounces sold decreased in Baltimore during the intervention period compared with the baseline year. In Washington, on the other hand, sales of recommended ounces increased over the course of the intervention. During the second year, adjusted sales increased almost 8% more in Washington than in Baltimore, a result that approached statistical significance. An increase in percentage of recommended sales over baseline was also seen in both years in Washington, compared with only year 1 in Baltimore. After adjustment, Washington showed a statistically significant relative increase in market share over Baltimore (2.4%) during the intervention period.

Frozen vegetables. The frozen vegetable category included most types of frozen vegetables; more than 60% of sales were in the recommended subcategory. Average sales changed little over the study period, although there was a small decline in percentage of recommended items sold. The decline in sales of recommended frozen vegetables was significantly smaller in Washington than in Baltimore, indicating a positive intervention effect.

Canned vegetables. The canned vegetable category included all canned vegetables. The recommended subcategory included corn, green peas, and lima beans. The number of ounces of vegetables in the recommended category fell during both study years compared with the baseline period, but the decline was greater in Baltimore than in Washington, resulting in a statistically significant increase in sales of the higher-fiber vegetables recommended by Eat for Health.

Canned and frozen beans. The canned and frozen bean category included baked and other canned beans and frozen beans, such as kidney beans and black-eyed peas; all items in this category were recommended. Intervention effects were negative but not statistically significant.

Dried beans. All types of dried beans were included in the dried bean category, and practically all were in the recommended subcategory. Sales declined from baseline in both areas, but again, the decline was smaller in Washington than in Baltimore, and positive intervention effects were seen for both intervention years. These changes approached statistical significance.

Dried fruit. The dried fruit category included raisins, dates, and other types of dried fruit; all were in the recommended subcategory. Results for this category were similar to those for dried beans in that intervention effects were positive but did not reach statistical significance.

Meat and poultry. The results of the separate analyses of the second-year meat and poultry interventions reveal limited changes, primarily associated with price, in the purchase of recommended vs other than recommended types during each 3-month intervention period related to the previous 1-year period. For meat, mean recommended ounces and percent recommended ounces decreased in both Baltimore and Washington; the decrease was smaller in Washington. For poultry, recommended ounces increased in both areas, but more so in Washington.

Discussion

Three goals were set for the Eat for Health program: (1) to demonstrate that a multifaceted nutrition information program could be carried out in more than 100 supermarkets over a period of 2 years; (2) to effect changes in food purchasing behavior in line with NCI's nutrition and cancer risk reduction dietary guidelines; and (3) to stimulate positive changes in knowledge and attitudes toward diet and health issues. Judged against these goals, Eat for Health had mixed success.

The ability to carry out such a large-scale, long-term program was strongly demonstrated. Over its 2-year span, from 150 000 to 200 000 bulletins were distributed each month for 25 months, more than 100 000 food guides were sold, and hundreds of thousands of people were exposed to a multimedia advertising campaign. Numerous groups and individuals were cooperatively involved in the preparation, review, and implementation of the many project elements. In the course of developing the program, Eat for Health staff were conscious of elements of previous supermarket interventions that had contributed to their successful implementation.[1,2,6,7,9] A number of these were found to work well for Eat for Health also. Important among them was the good working relationship between the Eat for Health program staffs at NCI and Giant Food. This relationship was formalized by an agreement outlining roles and responsibilities of both collaborators and further enhanced by the continual support of senior management at Giant and NCI. Training of Giant store managers and other staff and open discussion between store managers and the NCI staff person during monitoring visits also stimulated commitment to the program.

With respect to the second goal, Eat for Health's effect on food purchasing behavior appears to have been modest. This conclusion recognizes several limitations imposed by the complexity of the technical problems inherent in the analyses and by the realities of conducting research in a marketplace setting.

Because the intervention and control groups had to be assigned to Washington and Baltimore groups of stores rather than on a random basis, it was difficult to ensure comparability of the two groups and therefore to determine whether changes in sales were due to the intervention rather than to inherent differences in the study groups. Very large fluctuations in the levels of sales also occurred during the course of the intervention, resulting in tremendous variability in the data, which made it difficult to detect any intervention effect.

Despite these limitations, some changes in food purchasing behavior were observed, particularly with respect to produce. This may have been due to the emphasis on fruits and vegetables in the monthly bulletins and produce signs, to the popular weekly half-price sales in which the Giant salad bar was often fea-

tured, and to a public perception of the superiority of fresh produce.

Results in the third area—nutrition knowledge and attitudes and self-reported food purchasing behavior—appear to have been the most modest. Several possible reasons for this slight effect can be suggested. Washington-area shoppers may have already been sensitized to the general diet and health messages of Eat for Health by the two previous nutrition interventions held in Giant Food stores (Foods for Health in 1979 and Special Diet Alert in 1981) and by Giant's well-known commitment to providing nutrition information in its stores even in the absence of a formal intervention. National public health and private sector media campaigns related to various food and health issues may also have previously affected the knowledge and behaviors of Washington-area consumers. Finally, intervention stores happened to be located in more affluent areas than were comparison stores, and this somewhat higher socioeconomic status may account in part for the greater awareness of Washington shoppers. It should be noted, however, that the greatest positive change occurred for knowledge of a relationship between diet and cancer. This was a relatively new piece of information for consumers and the fact that such a dramatic change occurred here argues in favor of the effectiveness of Eat for Health materials in transmitting concepts to consumers.

Like the sales data, these survey data were analyzed in light of several recognized concerns and limitations. First, there was concern about the reliability of consumer responses to nutrition surveys. In particular, we were sensitive to the experience of previous programs, such as the Special Diet Alert and the Lean Meats Make the Grade programs, in which survey respondents claimed awareness of the program even before its inception or demonstrated confusion between the program and other types of nutrition promotions in the store.[2,6] To some extent this was true for Eat for Health as well, although knowledge of Eat for Health per se increased dramatically over the study period. Second, as this was a quasi-experimental study, there was very little control over survey participants' shopping behavior and media access, and therefore the sample of shoppers for the survey may not have been truly representative. Third, the time frame for data collection in the first wave was unexpectedly delayed, and data collection was still ongoing for several days after the program was launched.

Conclusions

One way to encourage and strengthen positive behavior change is to institutionalize its catalysts. Providing appealing, usable information in the supermarket that complements the healthful choices available there is one example of this process. Supermarket programs have demonstrated their value in enhancing customers' satisfaction with and loyalty toward their stores.[19,20] As a result, nutrition programs appear to be an entrenched feature of the supermarket business landscape.[21,22]

As messages and support systems for healthful dietary change become more established in the marketplace and diffused throughout society, it may be increasingly difficult to detect an impact from any one supermarket intervention. On the other hand, the positive contributions of these programs—along with those of many other nutrition efforts—may be reflected over time in the gradual lowering of nutrition-related chronic disease rates. □

References

1. Ernst NE, Wu M, Frommer P, et al. Nutrition education at the point of purchase: the Foods for Health project evaluated. *Prev Med.* 1986;15:60–73.
2. Levy AS, Mathews O, Stephenson M, Tenney JE, Schucker RE. The impact of a nutrition information program on food purchases. *J Public Policy Marketing.* 1985;4: 1–13.
3. Russo JE, Staelin R, Nolan CA, Russell GJ, Metcalf BL. Nutrition information in the supermarket. *J Consumer Res.* 1986; 13:48–70
4. Jeffery RW, Pirie PW, Rosenthal BS, Gerber WM, Murray DM. Nutrition education in supermarkets: an unsuccessful attempt to influence knowledge and product sales. *J Behav Med.* 1982;5:189–200.
5. Soriano E, Dozier DM. Selling nutrition and heart-healthy behavior at the point-of-purchase. *J Appl Nutr.* 1978;30:56–65.
6. Mullis RM, Pirie P. Lean Meats Make the Grade—a collaborative nutrition intervention program. *J Am Diet Assoc.* 1988;88: 191–195.
7. Mullis RM, Hunt MK, Roster M, et al. The Shop Smart for Your Heart grocery program. *J Nutr Educ.* 1987;19:225–228.
8. Dougherty MF, Wittsten AB, Guarino MA. Promoting low-fat foods in the supermarket using various methods, including videocassettes. *J Am Diet Assoc.* 1990;90: 1106–1108.
9. Worsley A, Worsley A, McConnon S. Evaluation of the New Zealand Heart Food Festival 1988–89. *Health Promotion Int.* 1990;5:127–135.
10. Pennington JAT, Wisniowski LA, Logan GB. In-store nutrition information programs. *J Nutr Educ.* 1988;20:5–10.
11. Mayer JA, Dubbert PM, Elder JP. Promoting nutrition at the point of choice: a review. *Health Educ Q.* 1989;16:31–43.
12. Glanz K, Mullis RM. Environmental interventions to promote healthy eating: a review of models, programs, and evidence. *Health Educ Q.* 1988;15:395–415.
13. Light L, Portnoy B, Blair JE, et al. Nutrition education in supermarkets. *Fam Community Health.* 1989;12(1):43–52.
14. Butrum RR, Clifford CK, Lanza E. NCI dietary guidelines: rationale. *Am J Clin Nutr.* 1988;48:888–895.
15. Light L, Tenney J, Portnoy B, et al. Eat for Health: a nutrition and cancer control supermarket intervention. *Public Health Rep.* 1989;104:443–450.
16. Odenkirchen JC, Portnoy B, Blair J, Rodgers AB, Light L, Tenney J. In-store monitoring of a supermarket nutrition intervention. *Fam Community Health.* 1992;14:1–9.
17. Block G, Hartman AM, Dresser CM, et al. A data-based approach to diet questionnaire design and testing. *Am J Epidemiol.* 1986;124:453–469.
18. Patterson BP, Kessler LG, Wax Y, et al. Evaluation of a supermarket intervention: the NCI-Giant Food Eat for Health study. *Eval Rev.* 1992;16:464–490.
19. *Trends: Consumer Attitudes and the Supermarket.* Washington, DC: Food Marketing Institute; 1989.
20. *Trends: Consumer Attitudes and the Supermarket.* Washington, DC: Food Marketing Institute; 1991.
21. Lekosi L. Supermarkets shape up with new consumer programs. *Supermarket Business.* April 1989:29–34.
22. Duff M. Nutrition: new rules rally the healthful . . . and supermarkets respond. *Supermarket Business.* April 1989:21–28.

A Workplace Intervention for Increasing Outdoor Workers' Use of Solar Protection

Afaf Girgis, PhD, Rob W. Sanson-Fisher, PhD, and Alan Watson, MBBS, DDM

ABSTRACT

Objectives. Outdoor workers are at high risk of developing skin cancer. Primary prevention in this group can potentially reduce the incidence of skin cancer, and also potentiates the spontaneous remission of existing solar keratoses. A randomized controlled trial was conducted to evaluate a solar protection intervention targeting outdoor workers.

Methods. Outdoor workers were randomly allocated to an intervention (n = 65) or control group (n = 77). The intervention group received individual skin screening by a dermatologist and participated in an education session. Pre- and posttest outcome measures included solar protection behavior (assessed using a validated diary), knowledge, and attitudes.

Results. There was a significant increase (16%) in the percentage of outdoor workers who were using a high level of solar protection at posttest compared to pretest in the intervention group, but there was no change in the control group. Although both groups improved in their knowledge score, the intervention group showed a significantly greater improvement at posttest. No changes in attitudes were detected.

Conclusions. The findings suggest that changes in solar protection are achievable with outdoor workers. (*Am J Public Health.* 1994;84:77–81)

Introduction

Skin cancer is the most common form of cancer affecting Australians,[1] and Australia has the highest incidence of skin cancer in the world.[2] In 1987, 20 to 25 Australians per 100 000 were diagnosed as having cutaneous malignant melanoma,[3] and 823 Australians per 100 000 were diagnosed as having nonmelanocytic skin cancers.[4] In Australia, skin cancer results in 1000 deaths each year and has been estimated to cost the community $400 million annually.[5]

Although the burden of illness associated with skin cancer is significant, this disease is estimated to be almost 80% preventable[6] and almost 100% curable by the adoption of a correct and vigilant set of primary and secondary preventive behaviors.[7,8] Primary preventive behaviors, which limit exposure to ultraviolet radiation, not only offer the best prospect of reducing the incidence of skin cancer,[6,9] but also potentiate the spontaneous remission of existing solar keratoses[10] prior to possible malignant transformations.[11] A substantial reduction in the number of solar keratoses has been observed in outdoor workers who were able to reduce their sunlight exposure over a 12-month period.[10] Secondary preventive behaviors, or screening for skin cancers and seeking prompt medical advice, substantially improve the prognosis of existing carcinomas.[12]

There are particular groups, most notably outdoor workers, who are at increased risk of developing skin cancer, given their solar exposure. The work environment for this group intrinsically constitutes a health hazard,[12–14] and it has been found that those working in outdoor occupations are more susceptible to developing nonmelanocytic skin cancers.[15–18] Squamous-cell carcinomas and basal-cell carcinomas are reportedly almost twice as prevalent in long-term outdoor workers than in indoor workers, and outdoor workers are reportedly more than twice as likely to develop solar keratoses.[18] Outdoor workers are also more likely to develop malignant melanomas on parts of their body exposed to sunlight during working hours.[17,19,20] Hence, outdoor workers represent an important target group at risk of developing skin cancer.

While avoidance of sunlight exposure is the ideal,[6,9] such a prevention strategy cannot be exclusively incorporated into the work practices of outdoor employees. This population might, however, beneficially adopt several other primary protective initiatives, such as job planning for maximum utilization of shade areas, particularly around midday, and use of solar protection items such as protective clothing, broad-brimmed hats, sunscreens, and sunglasses. Along with such measures, which are supported by the Australian Council of Trade Unions Occupational Health and Safety Unit,[21] educational and behavioral interventions promoting the adoption of solar protection behaviors are also valued.

Afaf Girgis and Rob W. Sanson-Fisher are with the New South Wales Cancer Council Cancer Education Research Project, University of Newcastle, and Alan Watson is with the Department of Dermatology, Royal Newcastle Hospital, New South Wales, Australia.

Requests for reprints should be sent to The Secretary, New South Wales Cancer Council Cancer Education Research Project, Locked Mail Bag 10, Wallsend, New South Wales, 2287, Australia.

This paper was accepted July 21, 1993.

Note. Although this research was undertaken by the New South Wales Cancer Council Cancer Education Research Project Team, the views expressed here are the authors' and not necessarily those of the Cancer Council.

The aim of the present study is to evaluate the effectiveness of a workplace intervention in changing the knowledge, attitudes, and solar protection behaviors of outdoor workers, using a randomized, controlled trial design.

Methods

Procedure and Sample

This study involved outdoor workers from a local electrical supply authority employing approximately 1200 individuals in total. The types of tasks undertaken by these workers, which were generally performed in groups of two to five individuals, included digging trenches; lopping trees; erecting, treating, and maintaining electricity poles; and reading meters door to door.

In this randomized controlled trial, 12 depots containing outdoor workers were randomly allocated to a control or intervention group based on geographic location. This randomization procedure was used to minimize the risk of contamination between the two study groups. Employees were considered ineligible if they spent less than 80% of their day working outside or if they were going to be on leave during the study period. Following random allocation to groups, eligible employees were sent a letter from the chief investigator inviting them to participate in the research.

The effectiveness of the intervention was assessed by comparing pre- and posttest and between-group differences in solar protection behavior, knowledge, and attitudes. The pretest assessment of these variables was made in both groups during the week prior to the intervention. Following pretest measures, the intervention group participated in the intervention, which was conducted over a 1-week period. The posttest measures were collected 1 month following the completion of the pretest data. At pre- and posttest assessments, participants were asked to return their completed measures to the research team in sealed, prepaid, addressed envelopes to ensure confidentiality of information.

Pre- and Posttest Measures

Solar protection behavior diary. Consenting participants were required to complete a solar protection behavior diary for 5 consecutive working days. This diary was developed to gather self-report information about the prevailing weather conditions and about participants' clothing and sunscreen use on different parts of the body. This information allowed a protection score to be calculated for each participant for the 5-day period. The format of the diary was easy to follow, with responses being recorded by circling one of a number of options in each category. To encourage accurate recall, the diary for each day was divided into four 2-hour blocks of time between 7:00 AM and 3:00 PM, and participants were asked to complete each entry as soon as possible after the end of the 2-hour period rather than at the end of the day.

During the pretest phase, the accuracy of self-report by the target group was assessed by direct observation procedures. Both a member of the research team and a safety officer directly observed a proportion of the outdoor workers and completed the items in the diary corresponding to one of the 2-hour time periods. Kappa was used to assess the extent of agreement between the participants' responses and those made by direct observations, with significant kappa achieved in four of the five categories of interest: what was worn on the head ($\kappa = 0.71$, $P < .05$), on the face ($\kappa = 0.42$, $P < .05$), on the shoulders and arms ($\kappa = 0.64$, $P < .05$), and on the legs ($\kappa = 0.89$, $P < .05$). The only category in which disagreement was reported was what was worn on the feet, and this involved reporting whether long or short socks were worn with safety boots ($\kappa = -0.02$, $P > .05$). This difference had no bearing on the calculation of the final protection score. These results, therefore, indicate that the self-report solar protection behavior diary is a valid measure of solar protection behaviors in this target population.

Knowledge and attitudes. A questionnaire developed and previously pilot tested by the research team[22] was used in this study: part 1 comprised 6 items designed to gather information about the subjects' skin type and skin-checking behavior; part 2 consisted of 18 items that assessed knowledge about solar protection and skin cancer issues (maximum score = 23 points, including subitems); and part 3 consisted of 25 items designed to assess attitudes about solar protection and skin cancer. Responses to all items in the questionnaire were recorded by circling the appropriate response options. Details of age and gender were also obtained in the questionnaire, which demonstrated high levels of face and content validity.[23]

Principal components analysis using the statistical package BMDP, program P4M,[24] was used to identify factors from the 25 attitude items. Six subscales were extracted in this analysis, which together accounted for 41.1% of the variance. The four subscales that were internally consistent (Cronbach's alpha > .05[25]) measured participants' attitudes about their personal susceptibility to developing skin cancer (12.8%), risk-taking propensity (8%), perceived benefits of having a suntan (4.8%), and external locus of control (4.6%).

Intervention

The intervention consisted of two components: a skin screening session and an education session.

In the screening session, each participant had his or her skin individually checked by one of five practicing dermatologists. All screening was conducted in privacy, and the dermatologists were asked to rate the condition of the skin on two scales: a 7-point photodamage scale and an 8-point premalignant and malignant damage scale (Table 1). These scales were developed by the dermatologists and specialist surgeons for use in this study, based on evidence that accumulated sun damage to the skin is associated with the development of nonmelanocytic skin cancer.[17,26,27] Before the intervention was begun, a number of meetings were held to discuss the two skin scales and the extent and type of feedback to be given to the subjects. This was done to maximize the consistency among the five dermatologists in their assessment of sun damage and personal risk of the subjects. The dermatologists then informed each participant of his or her current level of skin damage and personal risk of developing more extensive damage and/or skin cancer. Where necessary, the dermatologists also advised the subjects to see their personal general practitioner for management of a lesion or for referral to a specialist for management.

In the education session, the New South Wales Cancer Council's Hunter Region education officer delivered a 30-minute lecture to the participants in groups of approximately 10 to 15 participants. The talk addressed skin cancer in Australia, the increased risk for outdoor workers, and the opportunities for prevention using both measures that are generally available and measures that are specifically issued in this workplace, such as sunglasses, wide-brimmed hats, wide brims and back-flaps for hard hats, sunscreens, long-sleeved shirts, and lightweight overalls. Participants were encouraged to ask questions, and pamphlets

supporting the lecture were also distributed.

Participants in the control group were offered the opportunity to take part in the screening and education intervention at the end of the posttest data collection.

Results

Study Sample

Of the initial sample of 127 employees in the intervention group, 14 were not eligible for participation in the study, 17 did not return any of the pretest data, and 10 completed only part of the pretest, leaving 86 employees with complete pretest data. A complete data set suitable for analysis of all variables was available for 65 (76%) of these individuals (64 men, 1 woman; mean age = 40.6 years, range = 22 to 63 years).

Of the initial sample of 136 outdoor employees in the control group, 5 were not eligible for participation in the study, 20 did not return any of the pretest data, and 13 completed only part of the pretest data, leaving 98 employees with complete pretest data. A complete data set suitable for analysis of all variables was available for 77 (79%) of these individuals (all men; mean age = 40.4 years, range = 23 to 61 years).

Analyses were conducted to assess whether there were any differences in baseline variables between participants who had completed only part of the study (e.g., pretest data only) versus those who had completed all parts. These comparisons revealed no significant differences in age ($P = .124$), baseline knowledge ($P = .395$), and baseline solar protection score ($P = .149$).

Photodamage and Premalignant and Malignant Damage

A test for trend was conducted on the control and intervention groups' results and indicated that the two groups did not differ from the linear trend on either the photodamage scale ($P = .444$) or the premalignant and malignant damage scale ($P = .571$). Given the small sample in a number of the cells, grades 4 to 6 were collapsed in the photodamage scale and grades 3 to 7 were collapsed in the premalignant and malignant damage scale, prior to conducting the test for trend. The proportions of outdoor workers with different levels of damage on the two scales are presented in Table 1. The two scales were found to be significantly correlated ($r = 0.57$, $P = .0001$), indicating a positive relationship between the degree of photodamage and premalignant/malignant change in the skin.

TABLE 1—Proportion of Outdoor Workers with Different Grades of Photodamage and of Premalignant and Malignant Damage to Skin

Grade of Damage	Outdoor Workers, %	
	Intervention Group (n = 86)	Control Group (n = 98)
Photodamage		
0 None	7.48	2.91
1 Minimal macular hyperpigmentation and fine wrinkling	22.43	18.45
2 Pigmentary changes more marked in association with mild yellowing and thickening and mild coarse wrinkling	31.78	39.81
3 Both hyper- and hypopigmentation with moderate degree of yellowing, coarse wrinkling, and fine wrinkling in one area	15.89	14.56
4 Involvements of more than one area to a grade of 3	16.82	23.30
5 Severe pigmentary change, laxity, fine and coarse wrinkling in one area	1.87	0.00
6 Changes of grade 5 in all areas	3.74	0.97
Premalignant and malignant damage		
0 None	42.99	30.10
1 Presence (or history of previous treatment) of solar keratoses in one area only	26.17	31.07
2 Presence (or history of previous treatment) of solar keratoses in two areas	16.82	21.36
3 BCCs but no solar keratoses	0.00	0.00
4 Presence (or history of previous treatment) of BCCs or SCCs as well as of solar keratoses	11.21	16.50
5 Severe dysplastic change (to include solar keratoses, BCCs, SCCs, or Bowen's disease) in one area	1.87	0.97
6 Total dysplastic change in one area	0.00	0.00
7 Total dysplastic change in all areas	0.00	0.00

Note. BCC = basal-cell carcinoma; SCC = squamous-cell carcinoma.

Solar Protection Behavior

A protection score was calculated for every opportunity available for protection. An opportunity was defined as the worker's being outdoors between 11:00 AM and 3:00 PM when there was no rain. Protection level was calculated by giving each participant an aggregate score for the use of solar protection measures on each of nine body regions. A body region was considered to be adequately protected if it was fully covered by clothing/hat or shaded at the time of the interview, and/or if sunscreen lotion with a sun protection factor of 15 or higher had been applied to that region, as recommended by Australian Cancer Councils.

The points assigned to each body region were weighted to reflect the comparative risk, as indicated in the literature, of that region developing melanoma or skin cancer.[13,28-30] Thus, the face, back, and legs were appointed a maximum protection score of 3 points each, the shoulders and upper arms were appointed a maximum protection score of 2 points each, and the remaining body areas (neck, lower arms, chest, and stomach) were appointed a maximum protection score of 1 point each. Using this relative weighting scale, participants' overall protection level was scored out of a possible maximum of 17 points. Participants who scored 13 or more points (>75% of the body protected) were classified as having high protection during the period of diary completion. This cutoff point was selected to ensure consistency with previous research.[31]

Using this definition, z scores[32] used to make between-group comparisons indicated no baseline difference between the two groups in the percentage of workers protected (intervention = 50%, control = 49%; $P = .904$) but a significant difference at posttest (intervention = 66%, control = 49%; $P = .042$), with a greater proportion of outdoor workers in the intervention group having high solar protection. McNemar's test[32] was used to assess the change in protection from pretest to posttest in each group. As Figure 1 illustrates,

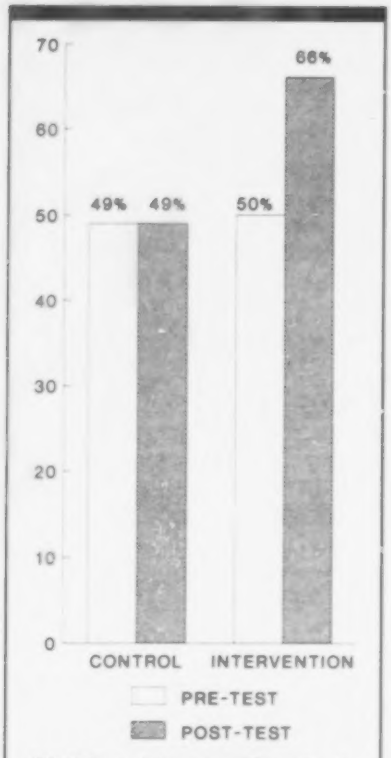


FIGURE 1—The proportion of outdoor workers (control vs intervention groups) using a high level of solar protection during the pre- and posttest periods.

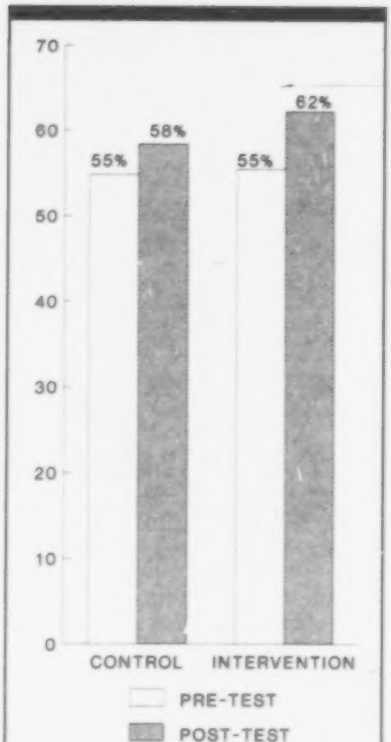


FIGURE 2—The percentage of correct knowledge score (control vs intervention groups) at the pre- and posttest periods.

the proportion of outdoor workers who used high solar protection increased significantly from pre- to posttest in the intervention group ($P < .02$) but did not change in the control group ($P = 1.0$).

Knowledge and Attitudes

A knowledge score was obtained for each group by calculating the average percentage of correct responses from a maximum of 23 points. Between-group knowledge was compared using *t* tests, and within-group knowledge was compared using paired *t* tests. As illustrated in Figure 2, there was no significant difference between the two groups at baseline (intervention = 55%, control = 55%; $P = .774$). While both groups demonstrated a significant improvement in the average percentage of correct knowledge score from pretest to posttest (intervention = $P < .001$; control = $P < .01$), the level at posttest was significantly higher in the intervention group (62% vs 58%, $P < .05$).

An attitude score was calculated by using the factor scores from each of the internally consistent factors, with the most desirable response receiving the maximum score (100%). There were no differences between the two groups either at baseline (intervention = 50%, control = 49%; $P = .567$) or at posttest (intervention = 49%, control = 48%; $P = .449$), and no changes were observed in either group from pretest to posttest.

Discussion

Although outdoor workers are at increased risk of developing skin cancer because of the nature of their occupation,[12–14] no evaluated health promotion activities directly targeting this health problem have been reported in the workplace. The primary aim of this study was to develop and evaluate the effectiveness of a workplace intervention program designed to improve the solar protection behavior and related knowledge and attitudes of a sample of outdoor workers in the Hunter Region.

The results revealed a significant improvement in both the knowledge and the solar protection behavior of the workers who participated in the intervention. This was indicated by an increase of 16% from pre- to posttest in the proportion of workers in the intervention group who used high solar protection, compared with no improvement in overall protection of workers in the control group. Knowledge about skin cancer and solar protection issues was also assessed in this study, and although both groups improved in their average knowledge score, the improvement was significantly greater in the intervention group. Neither group showed any change in their attitudes about skin cancer and solar protection.

The success of this intervention in improving the knowledge and solar protection behavior of the target group may be attributed to a combination of factors. First, the delivery of the intervention through the workplace may have contributed to its success, as previously discussed.[33–38] Given that exposure to the sun is an inherent part of the work environment of outdoor workers, attempts to change the workers' solar protection behavior at the work site are likely to be more successful than less environmentally specific interventions, as supported by the results of this study.

Second, the education session complemented the screening session by providing specific information on the means of modifying the health risk behavior (i.e., solar exposure), thereby addressing personal vulnerability to developing skin cancer. Such information included the opportunities to prevent skin cancer by using both measures that are generally available and measures that are specifically issued in this workplace, such as sunglasses, wide-brimmed hats, wide brims and backflaps for hard hats, sunscreens, long-sleeved shirts, and lightweight overalls. Furthermore, distribution of written information could encourage social support from other members of an individual's household, thereby providing an additional motivation to change.

In conclusion, although the intervention developed and evaluated in this controlled trial was associated with an improvement in both the level of knowledge and the proportion of workers adequately protected, a number of issues need to be addressed in future research. First, the proportion of protected outdoor workers improved 16% after participation in the intervention. While this result may not appear to be substantial, it was achieved by implementing a relatively brief intervention, which consisted of one screening/education session. Since the overall level of protected outdoor workers remains suboptimal in this group (66%), the poten-

tial for achieving greater improvements needs to be examined in future programs. Furthermore, research is required to develop and evaluate interventions that may have a lasting effect on the solar protection of the vulnerable target group.

Second, although the intervention was effective, further programs specifically targeting outdoor workers should be developed and systematically evaluated to assess the relative cost-effectiveness of simple versus more intensive interventions. Ideally, programs incorporating screening should use existing resources within the workplace, such as occupational health and safety officers, whose job description clearly encompasses such health promotion activities. The response of both employees and management in our research suggests that such programs would be well supported.

Finally, while this program focused specifically on individual behavioral change, it is important that the issue of solar protection for outdoor workers be considered on multiple levels. These may include providing educational sessions; adopting and enforcing workplace policies that address protection and screening; providing adequate solar protection items, including lightweight overalls, hats, and sunscreen; providing structural shelters, such as umbrellas or canopies; and restructuring outdoor work to avoid the most harmful period of the day. Each option is associated with some financial as well as administrative costs to the workplace, so the likelihood of a workplace adopting one or a number of these options will depend on available resources and the perceived priority of skin cancer in relation to other workplace health promotion and occupational health and safety issues. However, the long-term benefits in reducing the burden of illness of skin cancer are likely to be substantial. □

Acknowledgments

This research was undertaken by the New South Wales Cancer Council Cancer Education Research Project team, directed by R. Sanson-Fisher.

We gratefully acknowledge the Newcastle Area Dermatologists, including Drs A. Watson, B. Bartlett, R. Gillespe, J. Relic, J. Sippe, and G. Watt; Sisters S. Collins, D. Bradley, and D. Owens of the Newcastle Melanoma Unit; and Tony Tripodi for their involvement in the delivery of the intervention. We also thank the management, the safety officers, and the outdoor workers from Shortland County Council for their support and enthusiastic participation in the study. Finally, we acknowledge the extensive statistical assistance provided by Stephen Halpin and the research assistance provided by Angela Pirona-Barry.

References

1. Armstrong BK. The epidemiology and prevention of cancer in Australia. *Aust N Z J Surg.* 1988;58:179–187.
2. Marks R. Non-melanotic skin cancer and solar keratoses: the quiet 20th-century epidemic. *Int J Dermatol.* 1987;19:1045–1052.
3. Giles GG. *Mortality Rates in Australia 1985–1987.* Melbourne, Australia: Anti-Cancer Council of Victoria; 1989.
4. Giles GG, Marks R, Foley P. Incidence of non-melanotic skin cancer treated in Australia. *Br Med J.* 1988;296:13–19.
5. McCarthy WH, Shaw HM. Skin cancer in Australia. *Med J Aust.* 1989;150:469–470.
6. Stern RS, Weinstein MC, Baker SG. Risk reduction for non-melanoma skin cancer with childhood sunscreen use. *Arch Dermatol.* 1986;122:537–545.
7. Evans JF, Miller OF. Malignant melanoma and its precursors. *Postgrad Med.* 1986; 79(2):215–222.
8. Breslow A. Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. *Ann Surg.* 1970; 172:902–908.
9. Watson A. Sunscreen effectiveness: theoretical and practical considerations. *Aust J Dermatol.* 1983;24:17–22.
10. Marks R, Foley P, Goodman G, Hage BH, Selwood TS. Spontaneous remission of solar keratoses: the case for conservative management. *Br J Dermatol.* 1986;115: 649–655.
11. Bendl BJ, Graham JH. New concepts on the origin of squamous cell carcinomas of the skin: solar (senile) keratosis with squamous cell carcinoma—a clinicopathologic and histochemical study. *Proc Natl Cancer Conf.* 1971;6:471–480.
12. Sober AJ. Solar exposure in the aetiology of cutaneous melanoma. *Photodermatol.* 1987;4:21–23.
13. White AD. The use of sunscreens in Australia. *J Clin Med.* 1990;33(1):27–39.
14. Fitzpatrick TB, Sober AJ. Sunlight and skin cancer. *N Engl J Med.* 1985;313:818–819.
15. Marks R, Ponsford MW, Selwood TS, Goodman G, Mason G. Non-melanocytic skin cancer and solar keratoses in Victoria. *Med J Aust.* 1983;2:618–622.
16. Marks R, Selwood TS. Solar keratoses: the association with erythemal ultraviolet radiation in Australia. *Cancer.* 1985;56:2332–2336.
17. Beral V, Robinson N. The relationship of malignant melanoma, basal and squamous cell cancers to indoor and outdoor work. *Br J Cancer.* 1981;44:886–891.
18. Green A, Beardmore G, Hart V, et al. Skin cancer in a Queensland population. *J Am Acad Dermatol.* 1988;19:1045–1052.
19. Vagero D, Ringback G, Kiviranta H. Melanoma and other tumors of the skin among office, indoor and outdoor workers in Sweden. *Br J Cancer.* 1986;53:507–512.
20. Holman CDJ, Armstrong BK, Heenam PJ, et al. The causes of malignant melanoma: results from the West Australian Lions melanoma research project. In: Gallagher RP, ed. *Epidemiology of Malignant Melanoma.* Heidelberg, Germany: Springer Verlag; 1980.
21. ACTU Occupational Health and Safety Standard. In: Holmes N, Berger Y, eds. *Health and Safety Bulletin,* no. 61. Victoria, Australia: Australian Council of Trade Unions; 1989.
22. Foot G, Girgis A, Boyle K, Sanson-Fisher RW. Solar protection and related beliefs: a study of beachgoers. *Aust J Public Health.* 1993;17:209–214.
23. Bennett AE, Ritchie K. *Questionnaires in Medicine: A Guide to Their Design and Use.* London, England: Oxford University Press; 1975.
24. Dixon WJ. *BMDP Statistical Software.* Los Angeles, Calif: University of California Press; 1985.
25. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrica.* 1951;16:297–334.
26. Johnson BE. Solar radiation and skin cancer. *Br J Cancer.* 1973;28:91.
27. Holman CDJ, Armstrong BK, Evans PR, et al. Relationship of solar keratosis and history of skin cancer to objective measures of actinic skin damage. *Br J Dermatol.* 1984;110:129–138.
28. Balch CM, Soong SJ. Characteristics of melanoma that predict the risk of metastases. In: Costanzi JJ, ed. *Malignant Melanoma. Cancer Treatment and Research 9.* The Hague, the Netherlands: Martinus Nijhoff; 1983.
29. Giles C, Lecatsas S, Jolley D, Silver B. Trends in skin cancer in Victoria. *Trans Menzies Found.* 1989;15:153–155.
30. Ring IT, Walker SM, Firman D, et al. Melanoma in Queensland. *Trans Menzies Found.* 1989;15:157–163.
31. Girgis A, Sanson-Fisher RW, Tripodi AD, Golding T. Evaluation of interventions to promote solar protection in primary schools. *Health Educ Q.* 1993;20:275–288.
32. Bland M. *An Introduction to Medical Statistics.* Oxford, England: Oxford University Press; 1987.
33. Conrad P. Health and fitness at work: a participant's perspective. *Soc Sci Med.* 1988;26:545–550.
34. Zavella K, Davis L, Cottrell R, Smith W. Do only the healthy intend to participate in worksite health promotion? *Health Educ Q.* 1988;15:259–267.
35. Roman PM, Blum TC. Formal intervention in employee health: comparisons of the nature and structure of employee assistance programs and health promotion programs. *Soc Sci Med.* 1988;26:503–514.
36. Hollander RB, Lengerman JJ. Corporate characteristics and worksite health promotion programs: survey findings from Fortune 500 companies. *Soc Sci Med.* 1988; 26:491–501.
37. Taylor SE. *Health Psychology.* New York, NY: Random House; 1986.
38. Stachnik TJ, Stoffelmayr BE. Is there a future for smoking cessation programs? *J Health Educ.* 1982;7:47–56.

Racial Disparities in Reported Prenatal Care Advice from Health Care Providers

Michael D. Kogan, PhD, Milton Kotelchuck, PhD, MPH, Greg R. Alexander, ScD, MPH, and Wayne E. Johnson, PhD

ABSTRACT

Objectives. The relationship between certain maternal behaviors and adverse pregnancy outcomes has been well documented. One method to alter these behaviors is through the advice of women's health care providers. Advice from providers may be particularly important in minority populations, who have higher rates of infant mortality and prematurity. This study examines racial disparities according to women's self-report of advice received from health care providers during pregnancy in four areas: tobacco use, alcohol consumption, drug use, and breast-feeding.

Methods. Health care providers' advice to 8319 White non-Hispanic and Black women was obtained from the National Maternal and Infant Health Survey.

Results. After controlling for sociodemographic, utilization, and medical factors, Black women were more likely to report not receiving advice from their prenatal care providers about smoking cessation and alcohol use. The difference between Blacks and Whites also approached significance for breast-feeding. No overall difference was noted in advice regarding cessation of drug use, although there was a significant interaction between race and marital status.

Conclusions. These data suggest that Black women may be at greater risk for not receiving information that could reduce their chances of having an adverse pregnancy outcome. (*Am J Public Health.* 1994;84: 82–88)

Introduction

The relationship between maternal health risk behaviors during pregnancy (such as smoking, alcohol consumption, or illegal drug use) and adverse outcomes has been well documented. Maternal smoking has been associated with an increased risk of low birthweight, impaired fetal growth, fetal death, obstetric complications, and infant mortality.[1–4] Heavy alcohol consumption has been linked to a group of anomalies known as fetal alcohol syndrome.[5] Moderate or low alcohol consumption during pregnancy has been related to increased risks for preterm delivery, reduced birthweight, and spontaneous abortions.[6–8] Although it has been suggested that the evidence for linking moderate or low alcohol consumption with adverse outcomes is not conclusive,[9] health education messages continue to advise prudence in or abstinence from alcohol consumption during pregnancy. Illegal drug use, particularly use of cocaine or crack cocaine, has been associated with elevated risks for small-for-gestational-age births, premature births, abruptio placentae, and perinatal deaths.[10,11]

One way to alter these behaviors is through the advice and encouragement of women's health care providers. Most women are seen during the first trimester, when cessation of these behaviors could lower their risk of an adverse reproductive outcome.[12] As such, providers are in an advantageous position to identify pregnant women who are smoking, drinking, or using drugs and to initiate a health education program.[13] Studies have indicated that smokers are more likely to quit after receiving advice from a physician.[14,15]

Advice from providers may be particularly important in minority populations, who have higher rates of low-birthweight infants, premature births, fetal mortality, sudden infant death syndrome, and all-cause infant mortality.[16–18] There have been indications that Black women at high risk of giving birth to a low-birthweight infant may derive more important benefits from prenatal interventions.[19]

Analysis of racial disparities in prenatal care heretofore implicitly assumed that all prenatal care is the same. Yet the content of prenatal care may not be identical for all populations. The equivalency of the content of prenatal care has yet to be demonstrated, especially for all racial groups. Differential prenatal care may lead to differential efficacy and could be a factor in the large differential rates of birth outcomes by race seen in the United States.

Prenatal care interventions may be an important source of ameliorating racial disparities in maternal risk status and ultimately may be important for birth outcomes. A report by the US Public Health Service advocated examining the content of prenatal care.[20] Perceived maternal advice has not yet been extensively examined, especially by race.

Michael D. Kogan and Wayne E. Johnson are with the National Center for Health Statistics, Hyattsville, Md. Milton Kotelchuck is with the Department of Maternal and Child Health, University of North Carolina, Chapel Hill, NC. Greg R. Alexander is with the Maternal and Child Health Major, University of Minnesota, Minneapolis.

Requests for reprints should be sent to Michael D. Kogan, PhD, National Center for Health Statistics, 6525 Belcrest Rd, Room 840, Hyattsville, MD 20782.

This paper was accepted May 3, 1993.

Note. The opinions expressed in this paper are the authors' and do not necessarily reflect the views or policies of the institutions with which the authors are affiliated.

Editor's Note. See related editorial by Zapka (p 12) in this issue.

The objectives of this study were to (1) examine the percentages of Black and non-Hispanic White women who reported receiving advice from health care providers during pregnancy in four areas: tobacco use, alcohol consumption, drug use, and breast-feeding; and (2) determine whether any observed racial disparities were the result of other contributing factors.

Methods

These objectives were explored with data collected from the 1988 National Maternal and Infant Health Survey conducted by the National Center for Health Statistics. This was a follow-back survey consisting of three groups: 9953 women who had a live birth in 1988, 5332 women who suffered an infant death in 1988, and 3309 women who had a 1988 fetal loss. The survey was designed to be nationally representative and was drawn from the 1988 vital records of 48 states and the District of Columbia (South Dakota and Montana were not included). It included an oversampling of Blacks and low-birthweight infants. Approximately 50% of the respondents were Black, and 30% of the infants in the live birth sample had a low birthweight. Both married and unmarried women were included in the sample.

To adjust for this sampling frame, all live births were sorted into sampling strata formed by information taken from the birth certificate: mother's age and marital status and child's race and birthweight. To ensure an adequate sample size for analysis, different sampling rates were applied to each stratum. The same strata were used as nonresponse adjustment cells. The sample was then adjusted by post-stratification to once again be representative of the United States. A more complete description of the design of the 1988 National Maternal and Infant Health Survey has been published elsewhere.[21]

The National Maternal and Infant Health Survey used a mixed-mode methodology (mail, phone, or personal interview) to collect information from respondents. The response rate for women in the live birth cohort was 74.4%. With respect to birth certificate information, nonresponders were slightly more likely than responders to be Black and unmarried (data not shown).

This investigation included only women in the live birth cohort on the supposition that women with fetal deaths may not have been in care long enough to receive certain types of advice from their providers and that women with infant deaths may recall their prenatal care experiences differently. The study was also limited to White and Black women who reported receiving some prenatal care: 98% of the population. After other racial groups and women who classified themselves as Hispanic were excluded, 8310 women who had a live birth in 1988 were available for analysis. The study population is nationally representative for non-Hispanic White and Black live births only.

The outcome measures used in the present study included the mothers' responses (yes/no) to a series of questions regarding the receipt of any advice or instructions during any of their prenatal visits on (1) trying to breast-feed their baby; (2) reducing or eliminating consumption of alcohol; (3) reducing or eliminating use of tobacco; and (4) not using illegal drugs such as marijuana, cocaine, or crack cocaine.

Maternal race, education, household income, and marital status was determined from the mothers' responses to the questionnaire. Maternal age and the trimester that women began prenatal care was drawn from the birth certificate.

Respondents were asked where they went for most of their prenatal care. They were given a choice of private doctor's or nurse-midwife's office, county or city health department, community health center, health maintenance organization (HMO), work or school clinic, hospital clinic, hospital emergency room, or other site. In the analysis, county or city health department and community health center were combined into a variable called publicly funded sites of care. Work or school clinic and hospital emergency room were included in the "other" category because of small numbers (<1%).

Women were asked how their prenatal care was paid for. The choices were the respondent's or her partner's own income, insurance that the respondent carried or was carried for her, Medicaid, government assistance other than Medicaid (state or local), or other. Women were also asked whether they had received assistance from the Women, Infants, and Children's (WIC) program because women participating in the WIC program are supposed to receive advice on nutrition and health habits.

Information regarding the use of tobacco or alcoholic beverages in the 12 months before the respondents' infants were born was taken from the National Maternal and Infant Health Survey responses. Data on reported drug use were not included because of the known unreliability of such data. The National Maternal and Infant Health Survey did not inquire whether prior infants were breast-fed.

On the supposition that women who had a previous adverse outcome would be more likely to obtain or seek advice, a variable was created that categorized women into three risk groups based on their response to pregnancy history questions: women who did not report a previous pregnancy; women who had a previous pregnancy without an adverse outcome; and women whose previous pregnancy ended in either a stillbirth, miscarriage, abortion, or infant death.

Analysis

Because of the complex sampling method used in the National Maternal and Infant Health Survey, all analyses were weighted to be representative of the US national distribution for non-Hispanic Whites and Blacks.[21]

The data presentation is composed of three sections. First, demographics by race are presented descriptively. Second, the bivariate racial disparities are noted for the four outcome measures, as well as for the covariates. Third, logistic regressions were run to isolate the contributions of race.

Scaled weights were used to perform all analyses. The scaling factor was the reciprocal of the mean weight; the sum of all the scaled weights is the same as the actual number of observations.

The logistic regression analysis controlled for age, marital status, education, income, site of prenatal care, type of payment, maternal health behaviors, trimester that care began, and prior adverse pregnancy outcomes. In addition, interaction terms were examined in the multivariate models. The logistic regression analyses were conducted with the Survey Data Analysis software program.[22] This program was developed specifically for calculating variances in complex sample surveys. Adjusted odds ratios (ORs) and 95% confidence intervals were computed by using the beta coefficients and standard errors obtained from the logistic analyses.

Results

Table 1 shows the characteristics, by race, of the sample of 8310 women from the 1988 National Maternal and Infant Health Survey on all study sociodemographic, health system, health behaviors, and medical history variables, weighted to reflect their real population distributions.

TABLE 1—Characteristics of the Sample of 8310 Women from the 1988 National Maternal and Infant Health Survey, by Race

	Percentage of Weighted Sample[a]		
	White (n = 6782)	Black (n = 1532)	Total (n = 8310)
Race			
Black	...	...	18.4
White	...	...	81.6
Marital status			
Married	83.4	32.1	74.0
Divorced or separated	7.2	11.2	7.9
Single	9.4	56.6	18.1
Maternal education			
<12 y	12.1	27.0	14.8
High school graduate	40.3	43.1	40.8
>12 y	47.6	29.9	44.3
Maternal age, y			
15–19	9.4	22.6	11.8
20–29	59.2	58.6	59.1
30–34	23.2	13.5	21.4
35+	8.2	5.3	7.7
Income			
<$6 000	7.7	32.3	12.2
$6 000–$11 999	10.2	22.4	12.5
$12 000–$17 999	11.1	13.3	11.5
$18 000–$29 999	24.0	15.0	22.4
$30 000–$59 999	36.3	14.6	32.3
$60 000+	10.6	2.5	9.1
Prenatal care payment[b]			
Medicaid	12.5	47.7	18.9
Private insurance	68.2	31.2	61.4
Own money	38.8	15.3	34.5
Other government help	4.5	10.4	5.6
Other type of payment	6.8	8.4	7.1
Site of prenatal care			
Private office	73.5	39.9	67.4
Publicly funded site	8.5	26.5	11.8
Health maintenance organization	4.9	5.6	5.0
Hospital clinic	9.1	24.1	11.9
Other site of care	4.0	4.0	4.0
Received Women, Infants, and Children program assistance	20.0	58.2	27.0
Smoked in the year before delivery	34.1	25.3	32.4
Drank any alcohol in the year before delivery	55.4	26.3	50.0
Previous pregnancy history			
No previous pregnancy	35.4	36.5	35.6
Previous pregnancy with no adverse outcome	44.5	47.2	44.9
Previous pregnancy with adverse outcome	20.1	16.3	19.4
Trimester that prenatal care began			
First	85.1	64.1	81.4
Second	12.4	29.2	15.4
Third	2.5	6.6	3.2

[a]For all analyses, the figures given are the weighted-sample figures.
[b]Women could provide more than one source of payment.

Demographically, Black women giving birth in 1988 were distinct from White women giving birth in 1988. Black women were more frequently single, were less likely to be educated beyond high school, were younger, and had lower incomes. Black women also utilized publicly funded sites of care, the WIC program, and Medicaid programs more frequently than White women. Black women also reported better smoking and drinking health behaviors than White women.

Table 2 shows the bivariate association of race and all study covariates with the four health behavior advice variables. In all four areas, substantial numbers of women did not report receiving health behavior advice. Smoking cessation was the most common advice reported (69.5%), closely followed by cessation of alcohol consumption (68.4%). Receiving breast-feeding advice was reported by only 51% of all women in the United States.

White women reported receiving more prenatal advice on alcohol, smoking, and breast-feeding than did Black women. The disparity was greatest for avoidance of alcohol: only 60% of Blacks reported that they received advice on alcohol avoidance from their prenatal care provider compared with 70% of Whites.

For cessation of alcohol consumption, advice was, in general, substantial for all subgroups, but significantly more frequent for women of higher socioeconomic status (e.g., more often married, more than 12 years of education, and more income). Advice increased with age, through 30–34 years. Not surprisingly, advice on cessation of alcohol consumption was highest for alcohol users (76.1%), but even in those cases, 23.9% of alcohol users did not get advice. Although not presented in Table 2, the racial disparity remained after controlling for drinking status: 76.8% of White women who said they had drunk some alcohol in the 12 months before their delivery reported that they received advice compared with 69.7% of Black drinkers. Disparities across education and income groups seemed slightly stronger than disparities by race.

Advice on smoking cessation appeared to follow a different trend than advice on cessation of alcohol consumption. Demographically, younger women and women with less than 12 years of education received more advice. Income and marital status were less significant. Hospital clinics and other sites of prenatal care were the most likely to give advice; private offices were the least likely. Smoking advice was given to 90.4% of smokers and 59.5% of nonsmokers. Smoking status was, by far, the strongest predictor of smoking advice. Again, though not shown in the data presented, the racial disparity remained after controlling for the behavior status: 91.0% of White women who smoked in the year before delivery reported receiving advice compared with 86.5% of Black women who smoked. Disparities across age, WIC status, and smoking status seemed stronger than disparities by race.

Racial disparities were not noted for advice on cessation of drug use in the bivariate analysis. In general, advice about

cessation of drug use followed a pattern similar to that of smoking advice, with women of poorer socioeconomic status receiving more advice. Advice on cessation of illegal drug use was significantly more frequent for single, less educated, younger, and poorer women. Public clinics gave more advice than private sources of care.

Advice promoting breast-feeding was the advice reported least often. In general, there was some tendency for women of higher socioeconomic status to get more breast-feeding advice. Breast-feeding advice was more frequent in Whites, married women, and women with more than 12 years of education; it was least frequent in the lowest-income women. Site of prenatal care presents a complex picture, with HMOs and publicly funded clinics the most frequent providers of breast-feeding information. WIC participants reported only a 54.7% rate of receiving breast-feeding advice from their health care providers.

Table 3 shows the unadjusted and adjusted ORs (controlling for all variables in the logistic model) for not reporting receipt of advice on each of the four health behaviors, by race. Before adjustment, Black women were significantly more likely to report not receiving advice on cessation of alcohol consumption, smoking cessation, and breast-feeding promotion. After adjustment, a significant racial disparity in advice for alcohol and smoking cessation still remained. Breast-feeding promotion just missed reaching significance and was similarly skewed towards more advice for White women. The unadjusted OR for race in the analysis of drug use cessation was 0.99. When race was analyzed with the covariates, before interaction terms were assessed, the adjusted OR became significant (1.28), indicating that racial disparities were masked in the bivariate analysis. However, there was a significant interaction between race and marital status: Black single women were 1.4 times more likely than White single women not to receive advice on drug use cessation, whereas there were no racial differences among married women.

Table 4 presents the full logistic analysis for each of the outcome variables. For advice on cessation of alcohol consumption, only six variables were significant: drinkers were more likely to be given advice; and older women (>35 years), women with less than 12 years of education, Black women, WIC nonparticipants, and women who began prenatal care after

TABLE 2—The Association of Prenatal Care Advice with Race and Selected Covariates among the Sample of 8310 Respondents in the 1988 National Maternal and Infant Health Survey

	% Reporting Advice Received about Alcohol	χ^2	% Reporting Advice Received about Smoking	χ^2	% Reporting Advice Received about Drugs	χ^2	% Reporting Advice Received about Breast-feeding	χ^2
Race								
Black	60.2		64.2		66.1		47.3	
White	70.3	59.11***	70.7	25.17***	65.8	0.03	52.4	12.80***
Marital status								
Married	69.8		69.3		64.6		53.0	
Divorced or separated	66.1		69.0		67.3		50.4	
Single	63.9	21.45***	70.4	0.76	70.4	18.96***	45.4	28.23***
Maternal education								
<12 y	60.3		73.2		67.7		47.0	
High school graduate	67.5		70.7		66.3		51.3	
>12 y	72.1	61.89***	67.2	19.63***	64.9	3.65	53.1	13.94**
Maternal age, y								
15–19	61.8		71.4		71.1		50.6	
20–29	69.8		71.4		67.9		52.0	
30–34	71.8		67.2		61.5		50.4	
35+	59.1	59.52***	58.2	53.07***	55.5	65.94***	51.4	1.67
Household income								
<$6 000	63.7		70.8		69.9		45.9	
$6 000–$11 999	63.1		70.2		67.8		50.9	
$12 000–$17 999	64.7		68.5		67.1		55.1	
$18 000–$29 999	69.0		71.1		64.6		53.6	
$30 000–$59 999	71.0		68.7		64.6		51.0	
$60 000+	76.4	61.52***	67.1	6.62	64.2	13.95*	51.4	21.62**
Prenatal care payment[a]								
Medicaid	62.0		71.0		70.0		49.9	
Private insurance	70.3		68.2		64.0		50.5	
Own money	70.0		68.9		64.8		53.6	
Other government help	69.1		74.6		73.2		52.8	
Other type of payment	69.2		75.9		71.7		53.3	
Site of prenatal care								
Private office	68.3		68.6		63.4		50.6	
Publicly funded site	66.0		69.9		71.7		56.2	
Health maintenance organization	72.3		70.0		67.7		56.7	
Hospital clinic	68.4		71.7		70.8		49.0	
Other site of care	73.2	8.93	76.0	10.84*	74.1	51.52***	53.4	18.05**
Trimester that prenatal care began								
First	70.0		69.6		65.8		52.3	
Second	62.5		69.9		66.1		48.1	
Third	64.3	30.63***	64.0	3.83	67.6	0.43	49.4	7.80*
Received Women, Infants, and Children program assistance	65.9		74.3		71.6		54.7	
Did not receive Women, Infants, and Children program assistance	69.4	9.08**	67.7	32.81***	63.8	44.69***	50.2	13.13**
Drank alcohol in year before delivery	76.1							
Did not drink alcohol in year before delivery	60.8	225.33***						
Smoked in year before delivery			90.4					
Did not smoke in year before delivery			59.5	822.27***				

[a]Women could provide more than one source of payment; therefore, no χ^2 value was calculated.
*$P < .05$; **$P < .01$; ***$P < .001$.

TABLE 3—Unadjusted and Adjusted Odds Ratios Comparing Blacks with Whites on Prenatal Care Advice in the 1988 National Maternal and Infant Health Survey

Outcome	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Did not receive drinking cessation advice	1.57 (1.41, 1.74)	1.29 (1.10, 1.51)
Did not receive smoking cessation advice	1.35 (1.21, 1.50)	1.20 (1.01, 1.39)
Did not receive advice on breast-feeding	1.22 (1.11, 1.35)	1.15 (0.99, 1.32)
Did not receive advice on illegal drug use	0.99 (0.89, 1.09)	1.28 (1.10, 1.48)[a] 1.01 (0.65, 1.57)[b]

Note. OR = odds ratio; CI = confidence interval.
[a]Model without interaction.
[b]Model with significant interaction.

the first trimester were all less likely to be given advice.

For advice on smoking cessation, there were seven significant factors. Smokers were substantially more likely than nonsmokers to receive advice. Older and separated women received less smoking cessation advice. Although the bivariate analysis indicated that income was not significant, the multivariate analysis showed that lower-income women with incomes of less than $6 000, $6 000–$11 999, and $12 000–$17 999 received less advice than upper-income women. WIC nonparticipants also reported less advice.

For breast-feeding promotion, there were six significant factors, and each was stronger than race. Single women, women with less than 12 years of education, women with the lowest income levels, and WIC nonparticipants received less advice promoting breast-feeding. Women who received most of their prenatal care at publicly funded sites or HMOs or who had no private insurance were more likely to report receiving advice than women who received care at private physicians' offices.

Three factors predicted not receiving drug cessation advice. Race, in the presence of interaction, was not significant. Older women (ages 30–34 and 35+ years) and WIC nonparticipants received less advice; women who used either public prenatal care sites or hospital clinics received more advice than those who received care at private offices. Once again, WIC nonparticipants received less advice.

Interaction terms (with race) were examined for each of the four health behavior outcome measures. They were not significant or informative for smoking, alcohol, or breast-feeding advice. A significant interaction term (race by marital status) was noted for illegal drug use (OR = 1.43).

Discussion

Advice about prenatal health behavior is not a uniform feature of all prenatal care. Regardless of race, one third or more of the women surveyed reported receiving no prenatal advice on alcohol, tobacco, or drug use, and approximately 50% received no prenatal information on breast-feeding. The observation that women who smoked or drank were more likely to report receiving prenatal advice on tobacco or alcohol use is a positive indication that services were being targeted to at-risk groups. Notwithstanding, given the emphasis placed on the importance of providing all women with prenatal advice on substance use and breast-feeding, these findings indicate that much improvement is still needed in the content of prenatal care being provided to women in the United States.

The content of prenatal care is not uniform across racial groups. Compared with White women, Black women receiving prenatal care advice were significantly less likely to report receiving advice on drinking and smoking cessation, and the disparity in breast-feeding advice approached significance. This is the other critical finding of the study.

The current analyses suggest that although race is an important factor in the content of prenatal care, other programmatic and sociodemographic factors are equally, if not more, important. First, advice about two of the behaviors, smoking and drug use, was skewed towards poorer women, whereas advice about alcohol use and breast-feeding was skewed towards wealthier women. Health care providers may be giving advice based on their stereotypes of who is involved in what type of behaviors and not on a principal of equity. Second, the site of prenatal care was important. Advice on illegal drug use was more common for patients of publicly funded sites and hospital clinics than for private-office patients. Patients of HMOs and publicly funded sites were also found to have a lower risk of not receiving breast-feeding advice compared with private-office patients. Third, participation in the WIC program, which mandates prenatal care advice on these behaviors as part of its basic package of services,[23,24] had a protective effect in each multivariate analysis, with WIC nonparticipants reporting less prenatal advice. Fourth, older women (>35 years of age) were more likely to report not receiving advice on alcohol, tobacco, and drug use. This finding may reflect a perception on the part of the providers that these women were in less need of this advice because of earlier pregnancies, particularly in the case of illegal drug and alcohol use. Alternatively, providers may have perceived that these messages would be less effective in terms of changing established behaviors (e.g., tobacco use) and consequently may have stressed them to a lesser degree.

Although interactions were explored for each of the outcome measures, a significant interaction between marital status and race was only identified in the analysis of advice on illegal drug use. This finding suggests that White single women were targeted for advice on illegal drug use more often that Black single women or tended to report receiving such advice in greater proportions. These data are insufficient to propose an explanation for these findings. Illegal drug use is a sensitive area of discussion, and further investigation of these findings may need to explore to what extent differences in ethnic and cultural characteristics of providers and patients may inhibit the provision of advice in this area.

This study is limited in that it is based on the self-reports of the women surveyed. It is unclear whether women may be more likely to overestimate or underestimate the actual receipt of prenatal advice or whether error rates vary by type of advice, site of prenatal care, ethnicity of the mother, birth outcome, or other factors. Some studies have found that maternal recall is relatively accurate for birth outcomes,[25,26] whereas maternal recall of exposures during pregnancy has been mixed.[27,28] Moreover, patients and providers may have different recall on the content of a visit.[29] Nonetheless, it is women's perception, not the providers' report of their practice, that is ultimately most likely to be linked to health behavior changes.

Over the last two decades, considerable emphasis has been placed on the importance of adequate prenatal care for minority populations, who have been identified as having a greater risk of poor pregnancy outcome. To the extent that health care educational messages and campaigns have been effective, one might hypothesize that minority women who do receive prenatal care would be more likely than White women to be given information on these topics. These data indicate that this is not the case and suggest the need for continued and expanded medical education programs to increase provider awareness of the importance of these issues as part of prenatal care services to all women, particularly women at higher risk of poor pregnancy outcome.

The findings that there are variations in the content of prenatal care by ethnicity of the mother, site of care, and age, among other factors, also have implications for the interpretation of investigations focusing on the impact of the adequacy of prenatal care, as measured by the month of initiation of prenatal care and by the number of prenatal care visits. Indices of prenatal care utilization have been used to investigate ethnic differences in pregnancy outcome.[30,31] Although ethnic variations in prenatal care utilization have been repeatedly uncovered, the magnitude of these variations was insufficient to explain prevailing ethnic disparities in pregnancy outcome measures. The present findings, indicating that the content of prenatal care is not consistent for all ethnic groups, must now be considered as another potential explanation.

However, it should be stressed that although these data suggest that Black women receive less prenatal care advice on alcohol and tobacco use, it would be imprudent to overspeculate on whether these ethnic differences in the content of prenatal care advice are likely to appreciably explain the observed ethnic disparities in pregnancy outcome—given the lower likelihood of Black women's smoking and drinking before delivery. However, the content of prenatal care and the linkage of content and maternal needs in our understanding of the causes of racial disparities in birth outcomes must now be considered.

Conclusion

The present study suggests that large numbers of women of all races do not receive sufficient health behavior modification information as part of the content of their prenatal care. In particular, Black

TABLE 4—The Odds of Not Receiving Advice on Prenatal Care Behaviors during Pregnancy

	Alcohol Cessation OR (95% CI)	Smoking Cessation OR (95% CI)	Breast-feeding OR (95% CI)	Drug Use Cessation OR (95% CI)
Race: Black (vs White)	1.29 (1.10, 1.51)	1.20 (1.01, 1.39)	1.15 (0.99, 1.32)	1.01 (0.65, 1.57)
Marital status				
Single (vs married)	0.92 (0.75, 1.14)	1.02 (0.82, 1.26)	1.40 (1.15, 1.72)	0.76 (0.55, 1.05)
Separated or divorced (vs married)	1.01 (0.78, 1.32)	1.42 (1.07, 1.87)	1.15 (0.90, 1.46)	0.95 (0.73, 1.23)
Household income				
<$6 000 (vs ≥$18 000)	1.15 (0.91, 1.45)	1.30 (1.01, 1.66)	1.39 (1.11, 1.74)	1.07 (0.84, 1.36)
$6 000–$11 999 (vs ≥$18 000)	1.23 (0.98, 1.55)	1.39 (1.09, 1.77)	1.17 (0.94, 1.45)	1.11 (0.88, 1.40)
$12 000–$17 999 (vs ≥$18 000)	1.21 (0.97, 1.51)	1.40 (1.11, 1.77)	0.96 (0.78, 1.18)	1.06 (0.85, 1.32)
Maternal education				
<12 y (vs >12 y)	1.32 (1.06, 1.65)	1.21 (0.96, 1.54)	1.42 (1.15, 1.75)	1.16 (0.93, 1.45)
High school grad (vs >12 y)	1.13 (0.97, 1.33)	1.10 (0.94, 1.30)	1.13 (0.98, 1.30)	1.07 (0.92, 1.24)
Maternal age				
≤19 y (vs 20–29 y)	1.13 (0.90, 1.40)	0.93 (0.73, 1.18)	0.90 (0.73, 1.11)	0.91 (0.73, 1.15)
30–34 y (vs 20–29 y)	1.01 (0.85, 1.20)	1.13 (0.95, 1.36)	1.08 (0.92, 1.27)	1.27 (1.08, 1.50)
≥35 y (vs 20–29 y)	1.83 (1.43, 2.35)	1.61 (1.24, 2.08)	1.06 (0.83, 1.35)	1.64 (1.29, 2.09)
Women, Infants, and Children program participation (no vs yes)	1.28 (1.06, 1.55)	1.54 (1.26, 1.89)	1.57 (1.31, 1.88)	1.29 (1.00, 1.68)
Primary site of prenatal care				
Publicly funded site (vs private office)	0.87 (0.72, 1.06)	1.04 (0.77, 1.21)	0.79 (0.65, 0.97)	0.78 (0.62, 0.97)
Hospital clinic (vs private office)	0.84 (0.68, 1.03)	0.97 (0.69, 1.35)	1.05 (0.86, 1.29)	0.78 (0.63, 0.96)
Health maintenance organization (vs private office)	0.82 (0.60, 1.12)	0.86 (0.62, 1.20)	0.75 (0.57, 1.00)	0.75 (0.56, 1.02)
Other sites[a] (vs private office)	0.73 (0.50, 1.06)	0.69 (0.47, 1.00)	0.91 (0.66, 1.26)	0.65 (0.45, 0.94)
Type of payment for care				
Paid with own money (vs did not pay)	1.03 (0.87, 1.21)	1.04 (0.86, 1.24)	1.06 (0.91, 1.24)	1.06 (0.92, 1.27)
No insurance (vs private insurance)	1.00 (0.79, 1.25)	1.01 (0.80, 1.29)	0.75 (0.61, 0.93)	1.10 (0.86, 1.40)
Not paid by Medicaid (vs paid by Medicaid)	0.84 (0.63, 1.12)	0.89 (0.66, 1.19)	0.88 (0.68, 1.14)	1.12 (0.78, 1.60)
Not paid by other government program (vs paid by program)	1.09 (0.78, 1.54)	1.04 (0.72, 1.50)	0.95 (0.74, 1.25)	1.34 (0.95, 1.89)
Not paid by other source (vs paid)	1.07 (0.79, 1.46)	1.33 (0.96, 1.84)	0.96 (0.72, 1.28)	1.28 (0.93, 1.74)
Trimester that prenatal care began: second or third trimester (vs first)	1.13 (1.02, 1.25)	1.08 (0.97, 1.21)	1.06 (0.96, 1.17)	0.94 (0.85, 1.04)
Previous pregnancy history: prior adverse outcome (vs no pregnancy or no adverse outcome)	0.87 (0.73, 1.37)	0.98 (0.83, 1.17)	0.89 (0.75, 1.03)	0.83 (0.69, 1.01)
Drank alcohol in year before delivery (yes vs no)	0.53 (0.46, 0.61)			
Smoked in year before delivery (yes vs no)		0.15 (0.12, 0.18)		
Race by single marital status interaction				1.43 (1.00, 2.04)

Note. Odds are estimated from multiple logistic regression models among the sample of 8310 respondents in the 1988 National Maternal and Infant Health Survey. OR = odds ratio; CI = confidence interval.
[a]Other sites of care could be school clinic, work clinic, hospital emergency room, or other unspecified site.

women are more likely not to receive health behavior advice that could reduce their chances of having an adverse pregnancy outcome. Specifically, they are less likely to report receiving smoking and alcohol cessation advice. □

References

1. Kleinman JC, Madans J. The effects of maternal smoking, physical stature and educational attainment on the incidence of low birth weight. *Am J Epidemiol.* 1985;121: 843–855.
2. Kleinman JC, Pierre M, Madans J, Land GH, Schramm WF. The effects of maternal smoking on fetal and infant mortality. *Am J Epidemiol.* 1988;127:274–282.
3. Bergsjo P, Hoffman HJ, Davis RO, et al. Preliminary results from the collaborative Alabama and Scandinavian study of successive small-for-gestational age births. *Acta Obstet Gynecol Scand.* 1989;68:19–25.
4. *The Health Consequences of Smoking for Women. A Report of the Surgeon General.* Washington, DC: Public Health Service; 1980:189–239.
5. Jones KL, Smith DW. Recognition of the fetal alcohol syndrome in early infancy. *Lancet.* 1973;2:999–1001.
6. Berkowitz GS. An epidemiologic study of preterm delivery. *Am J Epidemiol.* 1981; 113:81–92.
7. Little RE. Moderate alcohol use during pregnancy and decreased infant birth weight. *Am J Public Health.* 1977;67:1154–1156.
8. Kline J, Shrout P, Stein Z, Susser M, Warburton D. Drinking during pregnancy and spontaneous abortions. *Lancet.* 1980;2: 176–180.
9. Walpole I, Zubrick S, Pontre J. Confounding variables in studying the effects of maternal alcohol consumption before and during pregnancy. *J Epidemiol Community Health.* 1989;43:153–161.
10. Handler A, Kistin N, Davis F, Ferre C. Cocaine use during pregnancy: perinatal outcomes. *Am J Epidemiol.* 1991;133:818–825.
11. Cherukuri R, Minkoff H, Feldman J, Parekh A, Glass L. A cohort study of alkaloidal cocaine ("crack") in pregnancy. *Obstet Gynecol.* 1988;72:147–151.
12. Sexton M, Hebel JR. A clinical trial of change in maternal smoking and its effect on birth weight. *JAMA.* 1984;251:911–915.
13. Stewart PJ, Dunkley GC. Smoking and health care patterns among pregnant women. *Can Med Assoc J.* 1985;133:989–994.
14. Ockene JK, Kristeller J, Goldberg R, et al. Increasing the efficacy of physician-delivered smoking interventions: a randomized clinical trial. *J Gen Intern Med.* 1991;6:1–8.
15. Glynn TJ, Manley MW, Pechacek TF. Physician-initiated smoking cessation program: the National Cancer Institute trials. In: Engstrom PF, Rimer B, Mortenson LE, eds. *Advances in Cancer Control.* New York, NY: John Wiley & Sons Inc; 1990: 11–26.
16. Kessel SS, Villar J, Berendes HW, Nugent RP. The changing pattern of low birth weight in the United States: 1970 to 1980. *JAMA.* 1984;251:1978–1982.
17. Shiono PH, Klebanoff MA. Ethnic differences in preterm and very preterm delivery. *Am J Public Health.* 1986;76:1317–1321.
18. Kessel SS, Kleinman JC, Koontz AM, Hogue CJ, Berendes HW. Racial differences in pregnancy outcome. *Clin Perinatol.* 1988;15:745–753.
19. Heins HC Jr, Webster NN, McCarthy BJ, Efrid CM. A randomized trial of nurse-midwifery prenatal care to reduce low birth weight. *Obstet Gynecol.* 1990;75:341–345.
20. *Caring for Our Future: The Content of Prenatal Care. A Report of the Public Health Service Expert Panel on the Content of Prenatal Care.* Washington, DC: Public Health Service; 1989.
21. Sanderson M, Placek PJ, Keppel KG. The 1988 National Maternal and Infant Health Survey: design, content, and data availability. *Birth.* 1991;18(1):26–32.
22. *Survey Data Analysis, Version 5.3.* Research Triangle Park, NC: Research Triangle Institute; 1990.
23. The Child Health and WIC Reauthorization Act of 1988, Pub L No. 100-147 (November 10, 1989). Amends Section 17 of the Child Nutrition Act of 1966 (42 USC § 1786).
24. The Anti-Drug Use Act of 1988, Pub L No. 100-690 (November 18, 1988).
25. Casey R, Rieckhoff M, Beebe SA, Pinto-Martin J. Obstetric and perinatal events: the accuracy of maternal report. *Clin Pediatr (Phila).* 1992;31(4):200–204.
26. Seidman DS, Slater PE, Ever-Hadani P, Gale R. Accuracy of mothers' recall of birthweight and gestational age. *Br J Obstet Gynaecol.* 1987;94:731–735.
27. Jacobson SW, Jacobson JL, Sokol RJ, Martier SS, Ager JW, Kaplan MG. Maternal recall of alcohol, cocaine, and marijuana use during pregnancy. *Neurotoxicol Teratol.* 1991;13:535–540.
28. Feldman Y, Koren G, Mattice K, Shear H, Pellingrini E, MacLeod SM. Determinants of recall and recall bias in studying drug and chemical exposure in pregnancy. *Teratology.* 1989;40:37–45.
29. Britt H, Harris M, Driver B, Bridges-Webb C, O'Toole B, Neary S. Reasons for encounter and diagnosed health problems: convergence between doctors and patients. *Fam Pract.* 1992;9:191–194.
30. Kessner DM, Singer J, Kalk CW, et al. Infant death: an analysis of maternal risk and health care. In: *Contrasts in Health Status, I.* Washington, DC: Institute of Medicine, National Academy of Sciences; 1973:1–132.
31. Alexander GR, Cornely DA. Racial disparities in pregnancy outcomes: the role of prenatal care utilization and maternal risk factors. *Am J Prev Med.* 1987;3:254–261.

Infant Feeding Policies in Maternity Wards and Their Effect on Breast-Feeding Success: An Analytical Overview

ABSTRACT

Objectives. The purpose of this review is to examine the plausibility of a causal relationship between maternity ward practices and lactation success.

Methods. Studies were located with MEDLINE, from our personal files, and by contacting researchers working in this field. Of the 65 studies originally reviewed, 18 met our inclusion criteria (i.e., hospital-based intervention, experimental design with randomization procedures, or quasi-experimental design with adequate documentation).

Results. Meta-analysis indicated that commercial discharge packs had an adverse effect on lactation performance. The impact of early mother-infant contact on lactation success was unclear. Rooming-in and breast-feeding guidance in a rooming-in context had a beneficial impact on breast-feeding among primiparae. Breast-feeding on demand was positively associated with lactation success. In-hospital formula supplementation of 48 mL per day was not associated with poor breast-feeding performance.

Conclusions. Hospital-based breast-feeding interventions can have a beneficial effect on lactation success, particularly among primiparous women. (*Am J Public Health.* 1994;84:89–97)

Rafael Pérez-Escamilla, PhD, Ernesto Pollitt, PhD, Bo Lönnerdal, PhD, and Kathryn G. Dewey, PhD

Introduction

The recent interest in maternity ward practices demonstrated by international health agencies[1] is justified. Rooming-in can have a beneficial impact on the health of the newborn,[2–4] and, as this review will show, there are strong reasons to believe that beneficial infant feeding policies in the maternity ward can contribute to lactation success. In 1989, the World Health Organization (WHO), in conjunction with the United Nations Children's Fund (UNICEF),[1] made several recommendations to promote breast-feeding in institutions that provide maternity services. These recommendations included rooming-in throughout the hospital stay, breast-feeding on demand, early initiation of breast-feeding, breast-feeding guidance by health personnel, and avoidance of supplementary fluids. Given the high number of births attended in clinics or hospitals worldwide,[4,5] an in-depth assessment of the impact of such policies is worthwhile.

In 1980, Winikoff and Baer[6] concluded that maternity ward policies similar to those advocated by WHO and UNICEF could have a positive impact on breast-feeding. However, they did not fully address the methodological limitations of the studies included in their review. Although Margen et al.[7] recognized that the majority of such studies did not account for confounders, they too concluded that lactation performance can be improved by encouraging mother-infant contact, providing breast-feeding information, and restricting bottle-feeding. Their conclusion may have been premature, however, since the only published study[8,9] on which they based their conclusion regarding long-term effects had serious methodological limitations.[10]

Bernard-Bonnin et al.[11,12] integrated the results of nine studies and used meta-analysis to test the effects of four different maternity ward policies on breast-feeding success. The authors concluded that lactation success was improved by both early mother-infant contact and breast-feeding support by health personnel with phone follow-up, but that in-hospital formula supplementation and breast-feeding support without phone follow-up had no significant impact. Their review did not consider policies such as discharge packs, rooming-in, and breast-feeding on demand, and it did not include any studies in developing countries.

Considering the limitations of these previous reviews, it is worthwhile to reexamine this issue. The intent of our review is to examine the plausibility of a causal relationship between maternity ward practices and lactation success, using the procedures of meta-analysis whenever possible.

Methods

This review is restricted to articles on the relationship between maternity ward practices and lactation success published in English or Spanish between 1951

Rafael Pérez-Escamilla, Bo Lönnerdal, and Kathryn G. Dewey are with the Department of Nutrition and the Program in International Nutrition, Dr Lönnerdal is also with the Department of Internal Medicine, and Ernesto Pollitt is with the Department of Pediatrics and the Program in International Nutrition, University of California at Davis.

Requests for reprints should be sent to Kathryn G. Dewey, PhD, Department of Nutrition, University of California at Davis, Davis, CA 95616-8669.

This paper was accepted September 1, 1993.

TABLE 1—Randomized Studies on the Effect of Commercial Discharge Packs on Lactation Performance

Authors	Reference No.	n	Intervention	Study Location	Population Characteristics
Bergevin et al.[a]	27	406	One 120-mL bottle One 235-mL ready-to-feed formula can One 120-g can formula powder Reusable plastic nipple Three pamphlets w/formula advertisements	Montreal, Canada	Well educated, urban, Caucasian, primiparae and multiparae, mean birthweight: 3.4 kg, mean gestational age: 39 weeks
Evans et al.[a]	28	95	One 4-oz ready-to-feed formula One 13-oz can concentrated formula One 16-oz can powdered formula	Seattle, Wash	Well educated, urban, Caucasian primiparae and multiparae[b]
Feinstein et al.[a]	29	166	*Commercial* Two water bottles Three breast-feeding pamphlets 32-oz can ready-to-feed formula *Research* Two water bottles Three breast-feeding pamphlets	Chicago, Ill	Urban Black primiparae and multiparae
Dungy et al.[c]	30	87	*Commercial* Infant formula *Research* Manual breast pump Breast pads Breast cream	Iowa	White, middle class, well educated[bd]
Frank et al.[a]	31	324	*Commercial* Two bottles of water Two nipples Health education pamphlets including formula advertisements *Research* Breast pads and breast-feeding pamphlets *without* formula advertisements	Boston, Mass	Low income, urban, Black, primiparae and multiparae, mean birthweight: 3.4 kg, mean gestational age: 39 weeks
Guthrie et al.[e]	32	134	One 454-g can of powdered infant formula	Cebu City, Philippines	Low income, urban, mean parity: 2.5–3.0[b]

[a]Interviewer(s) were blind to group assignment; subjects were unaware of study design.
[b]Mean birthweight and gestational age were not specified.
[c]Interviewer(s) were not blind to group assignment; not known if subjects were aware of study design.
[d]Parity was not specified.
[e]Interviewer(s) were blind to group assignment; subjects were aware of study design.

and 1991. Studies were located with MEDLINE, from our personal files, and by contacting researchers working in this field.

The initial screening criterion was that the studies should involve randomized trials or quasi-experimental designs. Of the 65 studies that were identified,[8,9,13–76] 35 met this criterion. Next, those studies with questionable internal validity (i.e., those that contained between-group socioeconomic differences [two studies],[24] self-selection of participants,[16,22,35] assignment to group based on physician's wishes,[20] and exceedingly high and unevenly distributed attrition rates[63] were eliminated. Also eliminated were those studies that did not provide enough information for their internal validity to be determined (i.e., they had unreported attrition,[40,64] poorly documented methodology,[21,37] or were unpublished brief communications[39,71]). Studies based on planned (vs practiced) breast-feeding behavior[46] or confounded by birth method (vaginal vs c-section)[65] or postdischarge breast-feeding interventions[66–68] were also excluded. In total, 12 out of 28 randomized studies and 5 out of 7 quasi-experimental studies were subsequently excluded.

Whenever possible, meta-analysis[77] was used to pool and interpret results across studies.

Results

The studies were grouped according to several of the policies in the WHO/UNICEF recommendations. The categories are commercial discharge packs, rooming-in and breast-feeding guidance, early maternal-infant contact, breast-feeding on demand, and in-hospital formula supplementation.

Commercial Discharge Packs

Following the 1981 WHO code for the marketing of breast milk substitutes,[78] five clinical trials were conducted in developed countries[27–31] and one in a developing nation[32] to test the effects of commercial discharge packs on lactation success (Table 1). Sample size ranged from 87 to 406 subjects. All the studies were conducted in urban areas, but there were differences between studies in the ethnic and socioeconomic characteristics of subjects.

In all studies, women were randomly assigned either to a group that received commercial discharge packs or to a control group. In half of the studies,[29–31] women in the control group received a research discharge pack (i.e., a pack with breast-feeding promotion materials); in the remaining three studies,[27,28,32] there was no intervention for controls. The con-

tent of the discharge packs is described in Table 1. We used meta-analysis[77] to examine the effect of commercial discharge packs on (1) any breast-feeding at 1 month, (2) full breast-feeding (i.e., breast milk as the only source of milk) at 1 month, and (3) any breast-feeding at 4 months. The results (Table 2, Figure 1) indicate that the distribution of commercial discharge packs had a detrimental effect ($P < .05$) on full breast-feeding at 1 month and on any breast-feeding at 4 months, and a marginal effect on any breast-feeding at 1 month.

The study conducted by Guthrie et al.[32] was the only one carried out in a developing country. Initially, the authors collected data for 8 months postpartum from poor women delivering in two public hospitals (hospitals A [n = 134] and B [n = 79]) in the Philippines. In each hospital, women were randomly assigned to a group receiving a free can of formula at hospital discharge or to a control group not given formula. In hospital A there was a trend toward lower breast-feeding rates among formula recipients, but in hospital B formula recipients tended to be more likely to breast-feed. However, the results from hospital B do not warrant conclusive inferences since the experimental group was very small (only 17 subjects at the end of the study). The authors "replicated" their original study in a third hospital and concluded that formula recipients were not less likely to breast-feed than controls. However, these women were followed for only 2 months rather than for 8 months.

We reanalyzed the data from hospital A using survival analysis[79] (instead of the chi-square analyses at different time points conducted in the original paper) and found that women who received the free can of formula were less likely to breast-feed ($P < .05$, one tail) during the first 8 months after delivery (Figure 2). This suggests that commercial discharge packs can have a negative impact on the lactation success of women in developing countries.

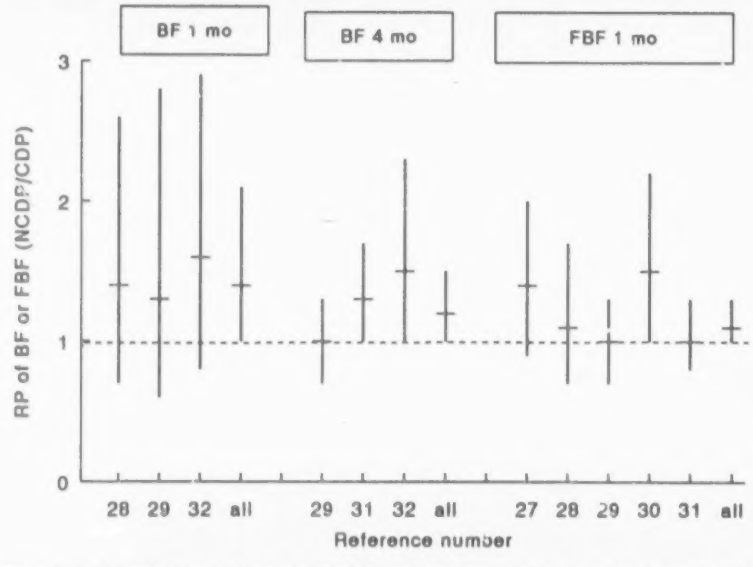
Another study that deserves further discussion is that of Bergevin et al.,[27] who documented that the negative impact of commercial discharge packs is more pronounced among vulnerable subgroups (e.g., primiparae, mothers with less education, and mothers who were ill after delivery). Evans et al.[28] could not replicate these findings; however, the sample sizes of the subgroups in their study were very small.

To summarize, commercial discharge packs are linked to poor lactation

TABLE 2—Homogeneity Test and Fisher's Combined Test of Studies Comparing Lactation Performance among Recipients (CDP) and Nonrecipients (NCDP) of Commercial Discharge Packs

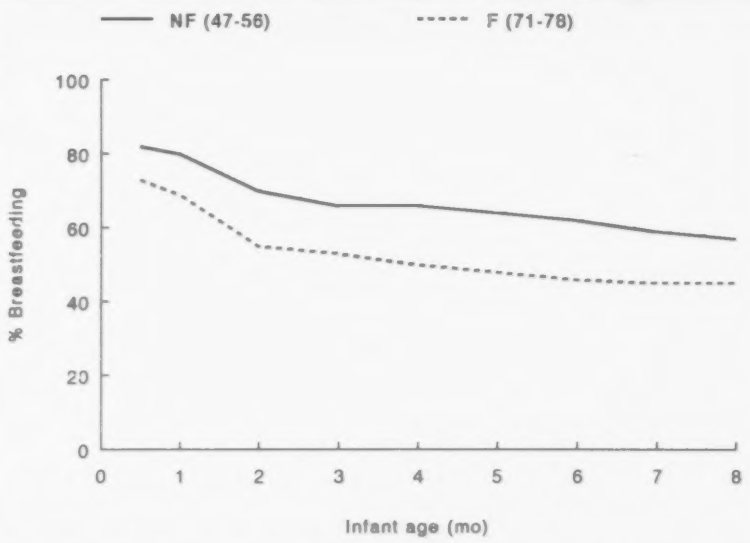
Study	CDP Breast-Feeding % (n)	NCDP Breast-Feeding % (n)	One-Tail P	Y[a]	Homogeneity Test[b]	"χ^{2}"	Fisher's Combined Test[d]
Any breast-feeding at 1 month							
Evans et al.[28]	65 (55)	75 (40)	.16	0.02	$\chi^2_{df=2} = 0.43$ ($P > .05$)	3.7	$\chi^2_{df=6} = 12.1$ ($.05 < P < .10$)
Feinstein et al.[29]	94 (76)	88 (90)	.25	0.23		2.8	
Guthrie et al.[32]	89 (78)	80 (56)	.06	0.18		5.6	
Full breast-feeding at 1 month							
Bergerin et al.[27]	78 (212)	84 (194)	.06	0.31	$\chi^2_{df=4} = 4.1$ ($P > .05$)	5.6	$\chi^2_{df=10} = 18.6$ ($P < .05$)
Evans et al.[28]	54 (55)	57 (40)	.39	0.04		1.9	
Feinstein et al.[29]	56 (76)	54 (90)	.54[f]	0.85		1.2	
Dungy et al.[30]	35 (44)	55 (43)	.02	2.49		7.8	
Frank et al.[31]	55 (167)	57 (157)	.34	0.40		2.1	
Any breast-feeding at 4 months							
Guthrie et al.[32]	50 (74)	67 (51)	.03	0.96	$\chi^2_{df=2} = 2.8$ ($P > .05$)	7.0	$\chi^2_{df=6} = 14.4$ ($P < .05$)
Feinstein et al.[29]	58 (76)	56 (90)	.82[f]	1.74		1.0	
Frank et al.[31]	55 (167)	65 (156)	.04	0.12		6.4	

[a] $Y = w(d - \bar{d})^2$, where $d = 2^*\text{ln}(1 - r^2)^{1/2}$, $r = (\chi^2/N)^{1/2}$, $w = 2^*N/(8 + d^2)$, $\bar{d} = \Sigma wd/\Sigma w$. N = study sample size, χ^2 = chi-square value.[77]
[b] Based on formula $\chi^2 = \Sigma w^* (d - \bar{d})^2$, with $K - 1$ degrees of freedom (where K = number of independent studies).[77]
[c] "χ^{2}" = $-2^* \log_e$ (one-tailed P).[77]
[d] $\chi^2 = \Sigma$"χ^{2}" with $2n$ degrees of freedom (where n = number of independent studies).[77]
[e] Breast-feeding 6 to 7 weeks after delivery.
[f] CDP group had higher breast-feeding rates than NCDP group. Therefore, $P_{1\text{tail}} = 1 - P_{2\text{tail}}/2$.[77]



Note. The horizontal lines indicate the RP of BF or FBF. The vertical lines depict the upper and lower bounds of the 95% confidence interval (CI) of the RP of BF or FBF (based on Mantel-Haenszel statistics). The RP of BF at 1 month with all studies combined (all) is 1.4 (95% CI:1.0, 2.1). The RP of BF at 4 months with all studies combined (all) is 1.2 (95% CI:1.0, 1.5). The RP of FBF at 1 month with all studies combined (all) is 1.1 (95% CI:1.0, 1.3).

FIGURE 1—Relative probability (RP) of postpartum breast-feeding (BF) or full breast-feeding (FBF) associated with receipt of commercial discharge packs (CDP) vs nonreceipt (NCDP), by study.



Note. F (n = 71–78) = can of powdered formula at discharge; NF (n = 47–56) = no formula provided at discharge. Based on data from Guthrie et al.[49]

FIGURE 2—Effect of formula samples at discharge on lactation success of Filipino women.

success, particularly among vulnerable subgroups such as primiparae and poor women in developing countries.

Rooming-In and Breast-Feeding Guidance

Only two studies that examined rooming-in and breast-feeding guidance[33,34] met our inclusion criteria. Since one of these two studies evaluated both rooming-in *and* breast-feeding guidance, these policies are discussed jointly in this section.

Lactation success of low-income primiparous women was studied in Managua, Nicaragua.[34] During the first 3 months of recruitment, mothers were randomly assigned to one of two groups: (1) complete maternal/infant separation throughout hospitalization with little or no breast-feeding promotion (n = 123), or (2) 45 minutes of maternal/infant contact immediately after birth followed by complete separation until discharge (n = 136). Women in the latter group received breast-feeding promotional messages. During the fourth and last month of recruitment, all eligible subjects were assigned to total rooming-in and to receive breast-feeding promotional messages (n = 116). Mothers and interviewers were blind to the study's hypotheses.

Groups were similar in several socioeconomic and demographic characteristics but differed in birthweight and obstetric procedures. None of the variables that differed was associated with lactation performance. At 1 week, the total rooming-in plus breast-feeding promotion group had a higher incidence ($P < .05$) of exclusive (63% vs 32%) and of any breast-feeding (93% vs 82%) when compared with the complete separation group. At 4 months, the differences between groups were not significant, but when the complete separation and the early contact plus breast-feeding promotion groups were combined, their breast-feeding rate at 4 months was lower ($P < .05$) than that of the total rooming-in plus breast-feeding promotion group (39% vs 50%). Thus, rooming-in accompanied by breast-feeding promotional messages had a positive impact on short- and longer-term breast-feeding success. However, this conclusion is tentative since, as recognized by the authors, the groups were not all randomly assigned, and between-group differences at baseline were identified. It was not possible to examine the effect of rooming-in by itself since breast-feeding messages were also provided to the total rooming-in group.

Pérez-Escamilla et al. conducted a study[33] among 165 low-income urban Mexican women who planned to partially or exclusively breast-feed. Women delivered in two different public hospitals. In one, infants were kept in a nursery (n = 58) and their diets were routinely supplemented with formula. In the other, rooming-in was the policy and formula supplementation was not allowed; in this hospital, women were randomly assigned to a group that received individual breast-feeding guidance during the hospital stay (n = 53) or to a control group in which normal routines were followed (n = 54). The groups were very similar in demographic and socioeconomic characteristics, prenatal care, breast-feeding motivation, previous breast-feeding exposure, maternal and infant anthropometry, and infant gender. None of the women mentioned infant feeding policies as a reason for choosing to deliver in a particular hospital.

Among primiparae, rooming-in (vs nursery) was associated with higher rates of full breast-feeding in the shorter but not in the longer term, whereas rooming-in plus breast-feeding guidance was associated with higher full breast-feeding rates throughout the 4-month study. These results suggest that rooming-in might not have a long-term effect unless it is accompanied by breast-feeding guidance. Rooming-in did not have any impact on lactation performance of multiparae. The quasi-experimental nature of this part of the study precludes causal inferences. However, the similarities in characteristics of the groups at baseline lend strength to the conclusions.

Among primiparae in the rooming-in hospital, the rate of decline in breast-feeding was significantly faster among those women who were only rooming-in than among those women who were rooming-in and receiving breast-feeding guidance. No significant differences between groups were found among multiparae. Because this part of the study included random assignment, it was concluded that the better lactation performance observed among primiparae in the rooming-in plus breast-feeding guidance group was indeed a result of the breast-feeding guidance intervention.

To summarize, the two studies that met our inclusion criteria suggest that both rooming-in and breast-feeding guidance in a rooming-in context can have a positive impact on lactation success among primiparae.

Early Maternal-Infant Contact

The handling of the newborn in most maternity wards involves routine procedures that cause maternal/infant separation soon after delivery. This separation may last a few (2 to 4) hours in wards with rooming-in or might continue until hospital discharge in wards with separate nurseries.[4,7]

TABLE 3—Randomized Studies on the Effect of Early Contact on Lactation Performance

Authors	Reference No.	n	Intervention	Country	Population Characteristics
Sosa et al.[a]	24	40	Initiated with nude infant on delivery table Duration: 45 min Dyad separated after contact and reunited 12–24 h postpartum	Guatemala	Urban poor primiparae Vaginal delivery Healthy term infant Birthweight: 2.04–3.85 kg
Ali and Lowry[b]	25	72	Skin-to-skin contact initiated shortly after delivery Contact duration: 45 min Dyad reunited 9 h after delivery	Jamaica	Urban poor primiparae and multiparae Vaginal delivery Full-term newborn Mean birthweight: 3.1 kg
Strachan-Lindenberg[b]	34	259	Initiated immediately after birth Duration: 45 min Dyad separated after contact and reunited until discharge Breast-feeding promotional messages	Nicaragua	Poor neighborhoods Primiparae Vaginal delivery Full-term newborn Mean birthweight: 3.1 kg
de Chateau et al.[a]	22, 23	40	Skin-to-skin contact initiated 10 min after delivery Duration: 10–15 min Midwife placed newborn at breast Husband present	Sweden	Middle-class primiparae Vaginal delivery Healthy term infant
Salariya et al.[c]	17	108	Initiated 10 min postpartum Midwife placed newborn at breast	United Kingdom	Primiparae "Normal healthy" infants Birthweight: 3.35 kg
Thomson et al.[b]	38	30	Skin-to-skin contact initiated 15–30 min postpartum Duration: 15–20 min Dyad separated after contact and reunited 12–24 h postpartum Partners welcomed	Canada	Primiparae Vaginal delivery Full-term newborn Birthweight > 2500 g Apgar ≥ 8
Taylor et al.[b]	36	50	Initiated 36–39 min postpartum Duration: 46 min Partners welcomed	United States	Vaginal delivery Full-term newborn Mean birthweight: 3.4 kg Mean Apgar: 8.3

[a]Not known if interviewer(s) were blind to group assignment and if subjects were unaware of study design.
[b]Interviewer(s) were blind to group assignment; subjects were unaware of study design.
[c]Mailed questionnaire; not known if subjects were unaware of study design.

Based on our initial screening criteria (i.e., randomized trials or quasi-experimental designs), we reviewed 14 studies[13,17,22,24(3 studies),25,34,35–40] on the effect of early contact on lactation success. Six of these studies were subsequently excluded[24(2 studies),35,37,39,40]; three of these[35,37,39] reported a positive impact of early contact on postpartum breast-feeding success whereas three[24,40] did not. Of the eight controlled trials that will be discussed, one[13] did not include a follow-up and therefore was not included in the meta-analysis. All except one of the trials in the meta-analysis[25] (Table 3) included only primiparae; three[24,25,34] were conducted in developing countries. Sample sizes ranged from 30 to 259. In all studies, women were randomly assigned to an "early contact" or a control group. In general, early contact meant that mother and newborn were allowed to have skin-to-skin contact beginning up to 39 minutes after delivery and lasting 10 to 45 minutes. In three studies,[22,34,38] mothers were encouraged to breast-feed during the period of early contact; there is no information on this issue in the remainder of the studies. It was reported that partners of the women were allowed to be present during early contact in three studies.[22,36,38] In four studies,[24,25,34,38] the infants were separated from their mothers after early contact and reunited later (within 9 to 24 hours, or at discharge). In the remaining studies, it is not clear if the newborns remained with their mothers after early contact.

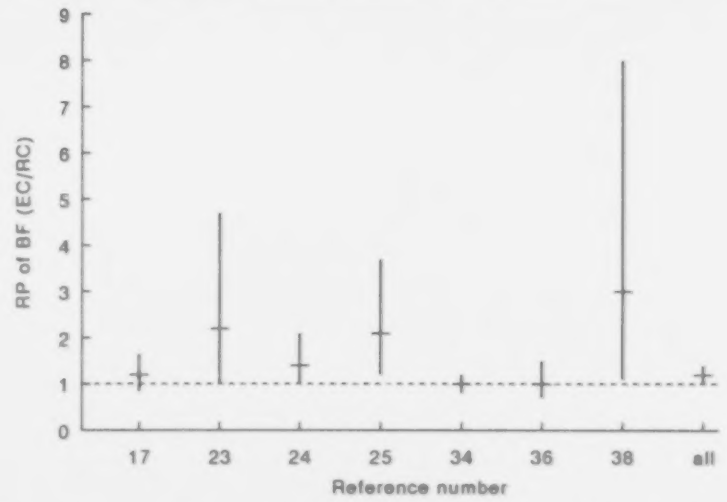
Meta-analysis indicated that early contact had a beneficial effect ($P < .05$) on the likelihood of breast-feeding at 2 to 3 months among primiparae (Table 4, Figure 3). However, these results should be interpreted with caution for two reasons. First, the effect size across studies was heterogeneous. This was mainly due to the study conducted in Nicaragua,[34] which found very similar breast-feeding rates among early and delayed contact groups. Second, it may not be possible to attribute improved breast-feeding rates directly to early contact since several studies also included breast-feeding guidance by health personnel or the presence of the father during early contact.

Woolridge et al.[13] conducted a clinical trial in Thailand to study the effect of early contact on the onset of milk production. Twelve dyads (i.e., mother-infant pairs) were randomly assigned to the comparison group, which included infants placed in a separate nursery and given water but not formula. About 24 hours after delivery, the newborns were returned to their mothers, who bedded-in throughout the hospital stay. Thirteen dyads were assigned to the early contact group, in which the newborns were brought to their mothers, fasted 2 to 4 hours after delivery, and remained with their mothers thereafter. The onset of milk production was assessed by test weighing a single feed each day on days 2 to 5 postpartum.

TABLE 4—Homogeneity Test and Fisher's Combined Test of Studies Comparing Lactation Performance at 2–3 Months among Regular Contact and Early Contact Groups

Study	Regular Contact % Breast-feeding (n)	Early Contact % Breast-feeding (n)	One-Tail *P*	Y[a]	Homogeneity Test[b]	$-\chi^{2}$[c]	Fisher's Combined Test[d]
Sosa et al.[24]	59% (20)	85% (20)	.04	1.2		6.4	
de Chateau et al.[23]	26% (19)	58% (21)	.02	1.8		7.8	
Strachan-Lindenberg[34,e]	51% (123)	50% (136)	.58	4.0	$\chi^2_{df=6} = 13.4$ ($P < .05$)[f]	1.1	$\chi^2_{df=14} = 40.9$ ($P < .05$)
Salariya et al.[17]	51% (53)	60% (55)	.17	0.04		3.5	
Thomson et al.[39,g]	20% (15)	60% (15)	.01	3.1		9.2	
Ali and Lowry[25,g]	27% (37)	57% (35)	.005	3.0		10.6	
Taylor et al.[38]	68% (19)	71% (31)	.31	0.3		2.3	

[a] $Y = w(d - \bar{d})^2$, where $d = 2^n r/(1 - r^2)^{1/2}$, $r = (\chi^2/N)^{1/2}$, $w = 2^n N(B + d^2)$, $\bar{d} = \Sigma wd/\Sigma w$. N = study sample size, χ^2 = chi-square value.[77]
[b] Based on formula $\chi^2 = \Sigma w^n (d - \bar{d})^2$ with $K - 1$ degrees of freedom (where K = number of independent studies).[77]
[c] $-\chi^{2n} = -2\Sigma \log_e$ (1-tailed P).[77]
[d] $\chi^2 = \Sigma\ -\chi^{2n}$ with $2n$ degrees of freedom (where n = number of independent studies).[77]
[e] Breast-feeding at 4 months.
[f] If reference 34 is excluded from meta-analysis, the effect size of remaining studies is homogeneous ($\chi^2_{df=5} = 6.3$ [$P > .05$]).
[g] Full breast-feeding at 2–3 months.



Note. The horizontal lines indicate the RP of BF. The vertical lines depict the upper and lower bounds of the 95% confidence interval (CI) of the RP of BF (based on Mantel-Haenszel statistics). The RP of BF at 2–3 months with all studies combined (all) is 1.2 (95% CI 1.0, 1.4).

FIGURE 3—Relative probability (RP) of breast-feeding (BF) 2–3 months after delivery associated with early contact (EC) vs regular contact (RC), by study.

The groups were similar in birthweight, parity, infant gender, and time of day at which milk intake was estimated on day 5. There was a difference of 21.5 hours in mean age at first contact and of 22.4 hours in mean age at first breast-feeding episode between the early contact and comparison groups, but there were no significant differences in milk production. However, the authors caution that the external validity of their findings may be limited since the study was conducted in a traditional society where the infants not only roomed-in but also bedded-in with their mothers. They also acknowledge that the effect of contact within the first 1 to 2 hours after delivery on the initiation of lactation was not studied.

Results presented in this section suggest that early maternal-infant contact might have a beneficial effect on lactation performance. However, to confirm this finding, additional studies with more rigorous methodology are needed.

Breast-Feeding on Demand

Three experimental studies[17,18,42] examined the impact of breast-feeding on demand.

Salariya et al.[17] studied the effect of both the timing of initiation of breast-feeding and the frequency of nursing during the hospital stay on the lactation success of 111 primiparous women. Women were stratified according to age and social class and were randomly assigned to follow one of four sets of feeding instructions: (1) 2E = early initiation (within 10 minutes after delivery), followed by nursing every 2 hours (n = 29); (2) 4E = early initiation, then nursing every 4 hours (n = 27); (3) 2L = later initiation (4 to 6 hours after delivery), followed by nursing every 2 hours (n = 27); and (4) 4L = later initiation, then nursing every 4 hours (n = 28). A questionnaire was mailed at 6 weeks, 12 weeks, and 18 months after delivery to assess infant feeding practices.

At 48 hours, the mothers who breast-fed every 2 hours (2E and 2L) were more likely to report onset of milk production than those who fed every 4 hours (4E and 4L) (54% vs 18%). Compared with infants fed every 4 hours, significantly fewer newborns fed every 2 hours required supplementary formula feeding (28% vs 9%). Median breast-feeding duration was longest in group 2E (182 days), followed by 4E (140 days), 2L (112 days), and 4L (77 days). It was concluded that both early initiation and increased frequency of breast-feeding improved lactation success. However, these results should be interpreted with caution because compliance by the mothers and blindness of interviewers to group assignment were not documented, the methodology used to evaluate the onset of milk production was not described, and no statistical tests comparing the median breast-feeding durations were reported.

De Carvalho et al.[18] conducted a study in the United States to investigate the effects of frequent and unrestricted breast-feeding on infants' milk intake and growth during the first month. Mothers delivering in July and August were told to feed their infants following the 3- to 4-hour hospital schedule (control group). Mothers delivering during September and October were encouraged to nurse on demand (experimental group). Breast-

feeding began within 6 hours after delivery. During the first 14 days, mothers recorded each breast-feeding. Milk intake was measured by test weighing on days 15 and 35. Initially, there were 47 control subjects and 28 experimental subjects. The dropout rates at 1 month were 66% and 57%, respectively.

Maternal age and parity were similar between groups, but mean birthweight was 478 g higher in the experimental group. Nursing frequency during the first 14 days was significantly higher in the experimental than in the control group (9.9 ± 1.9 vs 7.3 ± 1.4), but total nursing duration was similar (135 vs 138 min/24 h). Infants in the experimental group consumed significantly more milk on day 15 (725 vs 502 mL/24 h) and had gained significantly more weight since birth (561 vs 347 g) than those in the control group. At 35 days, nursing frequency was still significantly higher in the experimental group (9.8 ± 2.4 vs 6.8 ± 1.2), but the difference in total nursing duration was again not significant (177 vs 152 min/24 h). Infant milk intake on day 35 was marginally higher ($P = .08$) in the experimental group (841 vs 681 mL/24 h).

The authors concluded that frequent and unrestricted breast-feeding increases early milk production and infant weight gain. However, the between-group difference in birthweight is a potentially important confounding variable. Previous work by the same authors[41] indicated that birthweight was positively associated with nursing frequency and duration at 2 weeks, and others have shown that birthweight is correlated with breast milk intake.[40] In addition, the dropout rates were high and unevenly distributed across groups, and the potential self-selection bias caused by this was not assessed.

Illingworth et al.[42] conducted a study to determine the effect of in-hospital nursing on demand on the lactation performance of 241 mothers in Great Britain. All infants roomed with their mothers and in the first 2 days were breast-fed every 6 hours. On the third day, infants on one floor were fed every 4 hours (n = 106) while those on another floor were fed on demand (n = 131).

There were no significant between-group differences in parity or length of hospital stay. On the ninth day, significantly more newborns in the on-demand group had regained their birthweight as compared with the group fed on schedule (49% vs 36%). At 1 month, significantly more infants in the on-demand group were fully breast-fed (80% vs 64%). Moreover, significantly fewer mothers in that group had experienced sore nipples (13% vs 27%) or breast engorgement (17% vs 34%). Despite the intervention, there were no significant between-group differences in nursing frequency during the first week. However, it must be noted that nursing frequency for the group fed on schedule was assumed to be the six feedings that were scheduled.

In this study it is not known if the nurses, follow-up interviewers, and mothers were blind to the main goals of the study, and the assumption about group equivalence was based on two parameters only.

Although the three studies reviewed above all indicated a significant relationship between advice to breast-feed on demand and lactation success, serious methodological problems preclude any definitive conclusions.

In-Hospital Formula Supplementation

The practice of supplementing infants' diets with formula in the maternity ward is still pervasive. Only one study on this issue met our inclusion criteria.

Gray-Donald et al.[43] conducted a clinical trial in Montreal, Canada, to study the effect of in-hospital formula supplementation on the duration of breast-feeding. Subjects were assigned, based on bedspace availability, to one of two similar but separate maternity wards. In the formula supplementation group, the usual hospital routine was followed: all infants were given formula at 2 AM unless the mothers requested otherwise. Glucose water feedings were not restricted. The usual postpartum stay was 4 days. In the restricted formula nursery, mothers were routinely awakened at 2 AM to breast-feed; their infants were not supplemented with formula unless the mother requested it. Women were interviewed at 9 weeks postpartum by an interviewer blind to group assignment. Breast-feeding was defined as no more than one bottle-feeding per day.

The restricted formula (n = 388) and formula supplementation (n = 393) groups were very similar in birthweight and in socioeconomic and demographic characteristics. The percentage of infants unsupplemented in the hospital was significantly lower in the restricted formula than in the formula supplementation group (15% vs 63%), but the percentage receiving glucose water was similar. The percentage of mothers breast-feeding was very similar between groups at both 4 (71% vs 68%) and 9 weeks (55% vs 54%). No significant differences were observed even after reanalyzing the data using various breast-feeding definitions and within vulnerable subgroups.

The authors caution against generalizing from their findings. In-hospital supplementation was only 48 mL of formula per day. This contrasts sharply with the much higher amount of supplementation that has been reported in maternity wards in developing countries.[7]

Conclusions

This review presents strong evidence that several of the infant feeding policies recommended by WHO and UNICEF (i.e., discontinuation of commercial discharge packs, rooming-in, and breast-feeding guidance) can have a positive impact on lactation success. On the other hand, it also shows that the impact of other maternity ward policies (i.e., breast-feeding on demand) could not be properly evaluated because of serious methodological problems in the studies. Although our analysis suggests that early mother-infant contact might be related to lactation success, the meta-analysis was not easily interpretable due to study heterogeneity. Further clinical trials that control for confounders, such as husband's presence and breast-feeding guidance by health personnel, are required to confirm this relationship. The lack of effect of in-hospital formula supplementation on lactation performance of Canadian women[43] does not preclude the possibility that this practice has a negative impact on breast-feeding in settings where much higher levels of formula supplementation are common.[7] The potential differential impact of breast-feeding promotion efforts within vulnerable subgroups (e.g., low-income women and primiparae) deserves further consideration.

This review was restricted to infant feeding policies in maternity wards; we know almost nothing about the potential synergistic effect that these policies might have when combined with prenatal and postnatal breast-feeding interventions. It is important to fill in these gaps in knowledge in order to tailor cost-effective interventions whose aim is to increase the chances of lactation success and ultimately improve infant health. □

Acknowledgment

We are very grateful to Janet M. Peerson for her statistical advice and guidance with the meta-analyses.

References

1. *Protecting, Promoting and Supporting Breastfeeding: The Special Role of Maternity Services.* A joint World Health Organization/UNICEF statement. Geneva, Switzerland: World Health Organization; 1989.
2. Jelliffe DB, Jelliffe PEF. *Programmes to Promote Breastfeeding.* New York, NY: Oxford University Press; 1988.
3. Labbok M, McDonald M, eds. Proceedings of the interagency workshop on health care practices related to breast-feeding. *Int J Gynaecol Obstet.* 1990;31(suppl):1–191.
4. Brownlee A. Breastfeeding, weaning and nutrition: the behavioral issues. In: *Behavioral Issues in Child Survival Programs: A Synthesis of the Literature with Recommendations for Project Design and Implementation.* Washington, DC: US Agency for International Development; 1990:21–24, 55–64. Monograph 4.
5. UNICEF. *The State of the World's Children.* United Kingdom: P & L Adamson Press; 1992:85.
6. Winikoff B, Baer EC. The obstetrician's opportunity: translating "breast is best" from theory to practice. *Am J Obstet Gynecol.* 1980;138:105–115.
7. Margen S, Melnick V, Neuhauser L, Rios E. Associations between health facility practices and mothers' infant feeding practices after discharge. In: *Infant Feeding in Mexico: A Study of Health Facility and Mothers' Practices in Three Regions.* Emeryville, Calif: Printeam; 1991:26–40.
8. Mata L, Allen MA, Jiménez P, et al. Promotion of breastfeeding, health, and growth among hospital-born neonates, and among infants of a rural area of Costa Rica. In: Chen LC, Scrimshaw NS, eds. *Diarrhea and Malnutrition.* New York, NY: Plenum Press; 1983;12:177–202.
9. Mata L, Saenz P, Araya JR, et al. Promotion of breastfeeding in Costa Rica: the Puriscal study. In: Jelliffe DB, Jelliffe PEF, eds. *Programmes to Promote Breastfeeding.* New York, NY: Oxford University Press; 1988:55–69.
10. Pérez-Escamilla R. *Effect of the Maternity Ward System on the Lactation Performance of Low-Income Urban Mexican Women.* Davis, Calif: University of California; 1991. Doctoral diss.
11. Bernard-Bonnin AC, Stachtenko S, Girard G, Rousseau E. Hospital practices and breastfeeding duration: a meta-analysis of controlled trials. *Birth.* 1989;16:64–66.
12. Bernard-Bonnin AC, Stachtenko S, Rousseau E, Girard G. Pratiques hospitalières et durée de l'allaitement maternel: meta-analyse. *Rev Epidémiol Santé Publique.* 1989; 37:217–225.
13. Woolridge MW, Greasley V, Silpisornkosol S. The initiation of lactation: the effect of early versus delayed contact for suckling on milk intake in the first week post-partum. A study in Chiang Mai, Northern Thailand. *Early Hum Dev.* 1985;12:269–278.
14. Yamauchi Y, Yamanouchi I. The relationship between rooming-in/not rooming-in and breastfeeding variables. *Acta Paediatr Scand.* 1990;79:1017–1022.
15. Amatayakul K, Wongsawasdi L, Mungla-pruks A, et al. Successful sustained lactation following postpartum tubal ligation. *Adv Contracept.* 1991;7:363–370.
16. Mapata S, Djauhariah AM, Dasril D. A study comparing rooming-in with separate nursing. *Paediatr Indones.* 1988;28:116–123.
17. Salariya EM, Easton PM, Cater JI. Duration of breast-feeding after early initiation and frequent feeding. *Lancet.* 1978;2:1141–1143.
18. de Carvalho M, Robertson S, Friedman A, Klaus M. Effect of frequent breastfeeding on early milk production and infant weight gain. *Pediatrics.* 1983;72(3):307–311.
19. Dusitsin N, Chimpootaweep S, Taneypoon M, Boonsiri B. The effect of postpartum tubal ligation on breastfeeding. *J Thai Assoc Vol Steril.* 1979;1:51–57.
20. Herrera AJ. Supplemented versus unsupplemented breastfeeding. *Perinatol Neonatol.* 1984;8:70–71.
21. Schutzman DL, Hervada AR, Branca PA. Effect of water supplementation of full-term newborns on arrival of milk in the nursing mother. *Clin Pediatr.* 1986;25(2): 78–80.
22. de Château P, Holmberg H, Jakobsson K, Winberg J. A study of factors promoting and inhibiting lactation. *Dev Med Child Neurol.* 1977;19:575–584.
23. de Château P, Wiberg B. Long-term effect on mother-infant behaviour of extra contact during the first hour post partum: II. a follow-up at 3 months. *Acta Paediatr Scand.* 1977;66:145–151.
24. Sosa R, Kennell JH, Klaus M, Urrutia JJ. The effect of early mother-infant contact on breastfeeding, infection and growth. In: *Breastfeeding and the Mother.* Amsterdam, the Netherlands: Elsevier Publishing Co; 1976:179–193. Ciba Foundation Symposium, no. 45 (new series).
25. Ali Z, Lowry M. Early maternal-child contact: effects on later behavior. *Dev Med Child Neurol.* 1981;23:337–345.
26. Reiff MI, Essock-Vitale SM. Hospital influences on early infant feeding practices. *Pediatrics.* 1985;76:872–887.
27. Bergevin Y, Dougherty C, Kramer MS. Do infant formula samples shorten the duration of breastfeeding? *Lancet.* 1983;1: 1148–1151.
28. Evans CJ, Lyons NB, Killien MG. The effect of infant formula samples on breastfeeding practice. *J Obstet Gynecol Neonatal Nurs.* 1986;15:401–405.
29. Feinstein JM, Berkelhamer JE, Gruzka ME, Wong CA, Carey AE. Factors related to early termination of breast-feeding in an urban population. *Pediatrics.* 1986;78(2): 210–215.
30. Dungy C, Christensen-Szalanski J, Losch M, Russell D. Effect of discharge samples on duration of breastfeeding. *Pediatrics.* 1992;90:233–237.
31. Frank DA, Stephen JW, Sorenson JR, Heeren T. Commercial discharge packs and breastfeeding counseling: effects on infant-feeding practices in a randomized trial. *Pediatrics.* 1987;80(6):845–854.
32. Guthrie GM, Guthrie HA, Fernandez TL, Estrera NO. Infant formula samples and breastfeeding among Philippine urban poor. *Soc Sci Med.* 1985;20(7):713–717.
33. Pérez-Escamilla R, Segura-Millan S, Pollitt E, Dewey KG. Effect of the maternity ward system on the lactation success of low-income urban Mexican women. *Early Hum Dev.* 1992;31:25–40.
34. Strachan-Lindenberg C, Cabrera-Artola R, Jimenez V. The effect of early post-partum mother-infant contact and breast-feeding promotion on the incidence and continuation of breast-feeding. *Int J Nurs Stud.* 1990;27:179–186.
35. Taylor PM, Maloni JA, Brown DR. Early suckling and prolonged breastfeeding. *Am J Dis Child.* 1986;140:151–154.
36. Taylor PM, Maloni JA, Taylor FH. Extra early mother-infant contact and duration of breastfeeding. *Acta Paediatr Scand Suppl.* 1985;316:15–22.
37. Johnson NW. Breastfeeding at one hour of age. *Am J Maternal Nurs.* 1976;1:12–16.
38. Thomson ME, Hartsock TG, Larson C. The importance of immediate postnatal contact: its effect on breastfeeding. *Can Fam Physician.* 1979;25:1374–1378.
39. Sousa PLR, Barros FC, Gazale RV, et al. Attachment and lactation. Paper presented at the 14th International Congress of Pediatrics, Buenos Aires, Argentina; 1974.
40. Toussaint G, Casanueva E, Atkin L, Avila-Rosas H. La lactancia: impacto del encuentro temprano madre-hijo. *Perinat Reprod Hum (Mexico City).* 1988;2(3):115–120.
41. de Carvalho M, Robertson S, Merkatz R, Klaus M. Milk intake and frequency of feeding in breast-fed infants. *Early Hum Dev.* 1982;7:155–163.
42. Illingworth RS, Stone DHG, Jowett GH, Scott JF. Self-demand feeding in a maternity unit. *Lancet.* 1952;1:683–687.
43. Gray-Donald K, Kramer MS, Munday S, Leduc DG. Effect of formula supplementation in the hospital on the duration of breastfeeding: a controlled clinical trial. *Pediatrics.* 1985;75(3):514–518.
44. Griffin CCH, Popkin B, Spicer DS. Infant formula promotion and infant-feeding practices, Bicol Region, Philippines. *Am J Public Health.* 1984;74:992–997.
45. Winikoff B, Castle M, Laukaran V, eds. *Feeding Infants in Four Societies: Causes and Consequences of Mother's Choices.* Westport, Conn: Greenwood Press; 1989.
46. Procianoy R, Fernandes-Filho PH, Lazaro L, Sartori NC, Drebes S. The influence of rooming-in on breastfeeding. *J Trop Pediatr.* 1983;29:112–114.
47. Anand RK. A programme to promote breastfeeding in an urban hospital in India. In: Jelliffe DB, Jelliffe PEF, eds. *Programmes to Promote Breastfeeding.* New York, NY: Oxford University Press; 1988; 36–41.
48. Gainotti V, Pagani G. Promotion of breastfeeding in Santa Anna Hospital, Como, Italy. In: Jelliffe DB, Jelliffe PEF, eds. *Programmes to Promote Breastfeeding.* New York, NY: Oxford University Press; 1988; 42–47.
49. Verronen P. The breastfeeding promotion programmes at Tampere hospital, Finland. In: Jelliffe DB, Jelliffee PEF, eds. *Programmes to Promote Breastfeeding.* New York, NY: Oxford University Press; 1988: 30–35.
50. Winikoff B, Myers D, Laukaran VH, Stone R. Overcoming obstacles to breastfeeding in a large municipal hospital: applications

of lessons learned. *Pediatrics.* 1987;80: 423–433.
51. Gerstner GJ. Rooming-in: the Vienna experience. In: Jelliffe DB, Jelliffe PEF, eds. *Programmes to Promote Breastfeeding.* New York, NY: Oxford University Press; 1988: 55–71.
52. Gonzales RB. A large scale rooming-in program in a developing country: the Dr. Jose Fabella Memorial Hospital experience. *Int J Gynaecol Obstet.* 1990;31(suppl 1):31–34.
53. Jackson BE, Wilkin LC, Auerbach H. Statistical report on incidence and duration of breastfeeding in relation to personal-social and hospital maternity factors. *Pediatrics.* 1956;17(5):700–715.
54. Matulessy P. The Indonesian breastfeeding programme. In: Jelliffe DB, Jelliffe PEF, eds. *Programmes to Promote Breastfeeding.* New York, NY: Oxford University Press; 1988;166–172.
55. McBryde A. Compulsory rooming-in on the ward and private newborn service at Duke hospital. *JAMA.* 1951;145(9):625–628.
56. Relucio-Clavano N. The Baguio General Hospital and Medical Center breastfeeding and rooming-in programme. In: Jelliffe DB, Jelliffe PEF, eds. *Programmes to Promote Breastfeeding.* New York, NY: Oxford University Press; 1988:22–29.
57. Pérez-Escamilla R, Mejia LA, Dewey KG. Neonatal feeding patterns among low-income urban Mexican mothers. *Ecol Food Nutr.* 1992;27:91–102.
58. Loughlin HH, Clapp-Channing NE, Gehlbach SH, Pollard JC, McCutchen TM. Early termination of breastfeeding: Identifying those at risk. *Pediatrics.* 1985;75(3): 508–513.
59. Bloom K, Goldbloom RB, Robinson SC, Stevens FE. Factors affecting the continuance of breastfeeding: II. *Acta Paediatr Scand.* 1982;suppl 300:9–14.
60. Elander G, Lindberg T. Short mother-infant separation during first week of life influences the duration of breastfeeding. *Acta Paediatr Scand.* 1984;73:237–240.
61. Flores-Viladroza L, Brito-Duran E, Avila-Rosas H, et al. Exploración de factores que pueden modificar el establecimiento y la duración de la lactancia. *Perinat Reprod Hum (Mexico City).* 1988;2(1):16–24.
62. Helsing E, Kjaernes U. A silent revolution. Changes in maternity ward routines with regard to infant feeding in Norway 1973–1982. *Acta Paediatr Scand.* 1985;74:332–337.
63. Acosta Johnson C, Garza C, Nichols B. A teaching intervention to improve breastfeeding success. *J Nutr Educ.* 1984;16(1): 19–22.
64. Cohen SA. Post-partum teaching and the subsequent use of milk supplements. *Birth Fam J.* 1980;7:163–167.
65. Hall JM. Influencing breastfeeding success. *J Obstet Gynecol Nurs.* 1978;7:28–32.
66. Lynch AA, Koch MA, Hislop TG, Coldman AJ. Evaluating the effect of a breastfeeding consultant on the duration of breastfeeding. *Can J Public Health.* 1986; 77:190–195.
67. Jones DA, West RR. Lactation nurse increases duration of breastfeeding. *Arch Dis Child.* 1985;60:772–774.
68. Jones DA, West R. Effect of a lactation nurse on the success of breastfeeding: a randomised controlled trial. *J Epidemiol Community Health.* 1986;40:45–49.
69. Vandale-Toney S, Reyes-Vásquez H, Montaño-Uscanga A, López-Marroquín E, Vega-Castillo NE. Programa de promoción de la lactancia materna en el Hospital General de México: un estudio evaluativo. *Salud Pública Méx.* 1992;34:25–35.
70. Puga TF, Zaccaria A, Vukasovic JB, et al. Internación conjunta madre-hijo y lactancia materna. *Bol Méd Hosp Infant Méx.* 1979;36:1025–1050.
71. Slaven S, Harvey D. Unlimited suckling time improves breastfeeding. *Lancet.* 1981; i:392–393.
72. Houston MJ, Howie PW, Smart L, McArdle T, McNeilly AS. Factors affecting the duration of breastfeeding: part 2. early feeding practices and social class. *Early Hum Dev.* 1983;8:55–63.
73. Yamauchi Y, Yamanouchi I. Breastfeeding frequency during the first 24 hours after birth in full-term neonates. *Pediatrics.* 1990;86(2):171–175.
74. Kurinij N, Axelson ML, Forman MR, Weingold A. Predicting duration of breastfeeding in a group of urban primiparae. *Ecol Food Nutr.* 1984;15:281–291.
75. Kurinij N, Shiono PH, Rhoads GG. Breastfeeding incidence and duration in Black and White women. *Pediatrics.* 1988;81(3):365–371.
76. Tully J, Dewey KG. Private fears, global loss: a cross-cultural study of the insufficient milk syndrome. *Med Anthropol.* 1985;9:225–243.
77. Wolf FM. *Meta-Analysis: Quantitative Methods for Research Synthesis.* Beverly Hills, Calif: Sage; 1986.
78. *Marketing Code of Breast Milk Substitutes.* Geneva, Switzerland: World Health Organization; 1981.
79. *SAS/STAT User's Guide, Version 6.* 4th ed. Cary, NC: SAS Institute; 1990.
80. Dewey KG, Heinig J, Nommsen LA, Lönnerdal B. Maternal versus infant factors related to breast milk intake and residual milk volume: the DARLING study. *Pediatrics.* 1991;87:829–837.

Public Health Briefs

Correlates of High-Density Lipoprotein Cholesterol in Black and White Women

Gregory W. Heath, DSc, MPH, Caroline A. Macera, PhD, Janet B. Croft, PhD, Marcia L. Mace, MSPH, Thomas Gillette, MPH, and Frances C. Wheeler, PhD

ABSTRACT

The relationships of high-density lipoprotein (HDL) cholesterol with body composition, leisure-time physical activity, cigarette smoking, and education were examined in a community-based sample of 480 Black and 1337 White women. Univariate and multivariate analyses indicated inverse associations of HDL with body mass index and waist-to-hip ratio in both groups, and with cigarette smoking and low educational attainment among White women only. Since correlates of HDL cholesterol differ for Black and White women, further investigation of the differences in these correlates is warranted. (*Am J Public Health.* 1994;84:98–101)

Introduction

Low levels of high-density lipoprotein (HDL) cholesterol are independent predictors of coronary heart disease,[1] and high levels of HDL appear to be effective in preventing coronary heart disease in both Whites[2] and Blacks.[3,4] Although HDL levels are higher in women than in men,[5,6] very little information is available about the determinants or correlates of HDL in women, and especially Black women. Thus, associations of HDL with age, body mass index, waist-to-hip ratio, leisure-time physical activity, cigarette smoking, and educational attainment were examined in 1337 White women and 480 Black women.

Methods

Residents of two South Carolina towns were selected in 1987 by using random-digit dialing for households with telephones and simple random selection from city directories for households without telephones. Details of the study design, sampling procedures, interview response rates, examination procedures, and quality controls have been described previously.[7,8] Physical assessments in homes, work sites, and clinics were completed on 3202 (58%) of the 5489 men and women interviewed.

Blood samples were drawn on each subject using standardized protocols and were analyzed within 24 hours by the South Carolina State Laboratory, a participant in the Lipoprotein Standardization Program sponsored by the Centers for Disease Control and Prevention. Total cholesterol was analyzed using enzymatic techniques[9,10]; HDL levels were analyzed using the precipitation technique.[11]

The physical assessment examination included interviewer-measured height, weight, and circumferences of the waist and hip.[12] Body mass index was calculated as weight in kilograms divided by the square of height in meters; overweight was defined as a body mass index greater than 27.3 kg/m^2. Waist-to-hip ratio was calculated as waist girth divided by hip girth. Respondents who reported that they had not participated in any leisure-time physical activity during the previous month were defined as having engaged in no leisure-time physical activity. Education was categorized as less than high school graduate versus high school graduate or higher. A respondent who had smoked at least 100 cigarettes in his or her entire life and who currently smoked was defined as a cigarette smoker.

Descriptive statistics include ethnic-specific means and percentages of each variable, and Pearson correlation coeffi-

Gregory W. Heath is with the Division of Surveillance and Epidemiology, Epidemiology Program Office, Centers for Disease Control and Prevention, Atlanta, Ga. Carolina A. Macera is with the Center for Health Promotion and Disease Prevention, School of Public Health, University of South Carolina, Columbia. Janet B. Croft is with the Division of Chronic Disease Control and Community Intervention, National Center for Chronic Disease Prevention and Health Control, Centers for Disease Control and Prevention. Marcia L. Mace, Thomas Gillette, and Frances C. Wheeler are with the Center for Health Promotion, South Carolina Department of Health and Environmental Control, Columbia.

Requests for reprints should be sent to Gregory W. Heath, DSc, MPH, Division of Surveillance and Epidemiology, Epidemiology Program Office (C08), Centers for Disease Control and Prevention, 1600 Clifton Rd NE, Atlanta, GA 30333.

This paper was accepted May 11, 1993.

TABLE 1—Comparison of High-Density Lipoprotein (HDL) Cholesterol and Other Selected Characteristics of 1817 Women, by Race

	White (n = 1337)		Black (n = 480)	
	Mean (SD)	%	Mean (SD)	%
Age, y*	47.9 (16.4)	...	43.5 (17.3)	...
Total cholesterol	208.5 (46.9)	...	204.2 (43.4)	...
HDL*	49.7 (14.7)	...	54.2 (15.6)	...
Body mass index*	24.6 (5.2)	...	27.8 (6.8)	...
Waist-to-hip ratio*	0.81 (0.08)	...	0.84 (0.13)	...
Body mass index > 27.3*	...	22	...	45
No leisure-time physical activity*	...	34	...	61
Less than high school graduate*	...	23	...	39
Cigarette smoker	...	24	...	25

*$P < .001$ (*t* tests or chi-square tests for unadjusted differences between race groups).

TABLE 3—Age-Adjusted Mean High-Density Lipoprotein (HDL) Cholesterol in Women, by Race and Levels of Selected Characteristics

	White (n = 1337)		Black (n = 480)	
	Mean	Difference (95% CI)	Mean	Difference (95% CI)
Body mass index (kg/m^2)				
≤ 27.3	51.9	...	57.1	...
> 27.3	42.8	−9.1 (−10.9, −7.3)	50.9	−6.2 (−8.9, −3.3)
Any leisure-time physical activity	50.5	...	54.1	...
No leisure-time physical activity	48.2	−2.3 (−3.9, −0.6)	54.2	+0.1 (−2.8, +3.0)
High school graduate or higher	50.8	...	54.1	...
Less than high school graduate	46.2	−4.6 (−6.6, −2.6)	54.3	+0.2 (−3.1, +3.6)
Nonsmoker	50.5	...	54.3	...
Smoker	47.4	−3.1 (−5.0, −1.2)	53.8	−0.5 (−3.8, +2.7)

Note. CI = confidence interval.

TABLE 4—Multivariate Linear Regression Coefficients for the Association of Selected Characteristics with High-Density Lipoprotein (HDL) Cholesterol, for White and Black Women

	White (n = 1337)		Black (n = 480)	
	β	SE	β	SE
Age, y	.105*	.026	.032	.048
Body mass index	−.754*	.078	−.490*	.108
Waist-to-hip ratio (.01 unit)	−.190*	.051	−.128**	.058
No leisure-time physical activity	−.052	.832	.471	1.462
Less than high school graduate	−2.527**	.987	1.527	1.720
Smoker	−3.203*	.907	−1.206	1.641
r^2	.12		.07	

*$P \leq .01$, **$P \leq .05$.

cients that describe the unadjusted relationship of HDL with total cholesterol, age, body mass index, waist-to-hip ratio, leisure-time physical activity, smoking behavior, and educational attainment. Ethnic-specific multivariate linear regression techniques assessed the age-adjusted differences in mean HDL between levels of selected characteristics and the independent associations of age, body mass index, waist-to-hip ratio, leisure-time physical activity, educational attainment, and smoking behavior with HDL levels. All statistical procedures were implemented with SAS.[13]

TABLE 2—Correlations between High-Density Lipoprotein (HDL) Cholesterol and Independent Variables, by Race

	White (n = 1337)	Black (n = 480)
Cholesterol	.07**	.16***
Age, y	.03	.002
Body mass index	−.29***	−.22***
Waist-to-hip ratio	−.18***	−.13**
No leisure-time physical activity	−.07*	.003
Less than high school graduate	−.11***	.01
Smoker	−.09***	−.02

*$P \leq .05$; **$P \leq .01$; ***$P \leq .001$.

Results

Although they had a younger mean age than White women in this population, Black women had significantly higher ($P < .001$) levels of HDL, body mass index, and waist-to-hip ratio, and a greater proportion of them were overweight (Table 1). There were also significantly higher proportions of Black women reporting no leisure-time physical activity and low educational achievement than White women, but similar proportions of both groups were current smokers.

In both groups, HDL had a statistically significant inverse correlation with body mass index and waist-to-hip ratio, a significant direct correlation with total cholesterol, but no relation with age (Table 2). No leisure-time physical activity, less than 12 years of education, and current cigarette smoking were significantly inversely correlated with HDL in White women only. After controlling for age, statistically significant differences ($P < .05$) in mean HDL levels were observed between weight categories in both groups and between levels of leisure-time physical activity, education, and smoking among White women only (Table 3).

In multivariate linear regression, both body mass index and waist-to-hip ratio were independently associated with HDL among both groups (Table 4). In addition, cigarette smoking and low educational attainment were inversely associated and age was positively associated with HDL in White women only. Although antihypertensive medication can

affect certain lipid levels,[14] results were unchanged after excluding the 24% of Black women and 17% of White women who were taking such medicines.

Discussion

The purpose of this study was to examine the correlates of HDL in Black and White women. In the US population,[15] important correlates of higher levels of HDL include the female gender, African-American ancestry, and alcohol consumption. Smoking and body mass index had strong inverse relationships with HDL levels. Less strongly related were age, years of education, and high levels of physical activity. Sprafka et al.[16] also showed similar results in women with the use of exogenous hormones and beta-blockers identified as additional correlates. However, these studies did not include measures of body fat distribution, which have been shown in clinical studies to be important predictors of HDL.[17]

That body mass index had a significant inverse relationship to HDL in both groups of women in this and other studies is interesting, considering the much greater prevalence of overweight[18,19] yet higher levels of HDL[15] in Black women. Our results underscore possible biological and/or environmental differences between Black and White women, which would account for the observed differences in body composition, body weight, and body fat distribution. Body fat distribution is an independent predictor of risk for coronary heart disease.[3,20] As the most sensitive measure of body fat distribution,[21] waist-to-hip ratio has been shown to be a strong inverse correlate of HDL among older men[22] and young adults.[23] In the present investigation of women, waist-to-hip ratio correlated inversely with HDL.

In univariate analyses, lack of physical activity was inversely associated with HDL only in White women. Compared with White women, Black women were more likely to report no leisure-time physical activity. In the final multivariable models, lack of leisure-time physical activity failed to have any significant impact on HDL. This can be partially explained by the correlation between physical activity and measures of body fat and by the significant effects that body composition and body fat distribution have on lipoprotein lipid levels.[24] Cross-sectional data from Coon et al.[25] suggest that the best predictor of HDL is body composition and not physical activity per se or measured physical fitness. Therefore, our failure to see a persistent relationship between leisure-time physical activity and HDL is most likely owing to the overriding effects of body composition differences in our populations and especially in the Black women, a disproportionate number of whom had body mass index levels greater than 27.3 kg/m^2. Another possible explanation may be a lack of sufficient power because of the disproportionately smaller number of Black women compared with White women in our study.

In White women of the current study, plasma HDL increased with increasing educational attainment. In contrast, an inverse relationship between HDL and education that did not reach statistical significance was observed among Black women. This may also be owing to inadequate numbers of Black women in our study. However, these findings are similar to those found by Freedman et al.[26] and Wilson et al.[27] for Black men, whose lower educational attainment was associated with higher plasma levels of HDL. Our results support a role for educational attainment in determining risk status for coronary heart disease.[28]

In summary, this study demonstrates that the most powerful correlates of HDL for Black and White women are body size and body fat distribution, with cigarette smoking and educational attainment serving as moderate predictors of HDL for White women only. These analyses demonstrate the need for investigation of the possible biological and/or environmental differences between White and Black women with regard to the correlates of lipoprotein lipids. Also needed is exploration into the mediating role of body composition in determining lipoprotein lipid levels. □

Acknowledgments

This work was made possible by the provision of data from the South Carolina Department of Health and Environmental Control Cardiovascular Disease Prevention Project, which was supported by Cooperative Agreement U50/CCU 402234 from the Centers for Disease Control and Prevention.

References

1. Gordon DJ, Rifkind BM. High-density lipoprotein: the clinical implications of recent studies. *N Engl J Med.* 1989;321:1311–1316.
2. Manninen V, Mo E, Frick MH, et al. Lipid alterations and decline in the incidence of coronary heart disease in the Helsinki Heart Study. *JAMA.* 1988;260:641–651.
3. Cooper R, Sempos C, Ghali J, Ferlinz J. High-density lipoprotein cholesterol and angiographic coronary artery disease in Black patients. *Am Heart J.* 1985;110:1006–1011.
4. Freedman DS, Gruchow HW, Manley JC, et al. Black/White differences in risk factors for arteriographically documented coronary artery disease. *Am J Cardiol.* 1988;62:14–19.
5. Green MS, Heiss G, Rifkind BM, Cooper GR, Williams OD, Tyroler HA. The ratio of plasma high-density lipoprotein cholesterol to total and low-density lipoprotein cholesterol: age-related changes and race and sex differences in selected North American populations: the Lipid Research Clinics Program Prevalence Study. *Circulation.* 1985;72:93–104.
6. Van Horn LV, Ballew C, Liu K, et al. Diet, body size, and plasma lipids-lipoproteins in young adults: differences by race and sex. *Am J Epidemiol.* 1991;133:9–23.
7. Wheeler FC, Lackland DT, Mace ML, Reddick A, Hogelin G, Remington PL. Evaluating South Carolina's Community Cardiovascular Disease Prevention Project. *Public Health Rep.* 1991;106:536–542.
8. Macera CA, Heath GW, Eaker ED, Croft JB, Yeager KK, Wheeler FC. Leisure-time physical activity and high-density lipoprotein cholesterol in a biracial community sample. *Ethnicity Dis.* 1993;3:152–157.
9. *Manual of Laboratory Operation.* Bethesda, Md: National Institutes of Health, Lipid Research Clinics Program; 1974;1. DHEW publication NIH 75–628.
10. Siedel J, Hagele EO, Ziegenhorn J, Wahlefeld AW. Reagent for the enzymatic determination of serum total cholesterol with improved lipolytic efficiency. *Clin Chem.* 1983;29:1075–1080.
11. Kostren GM, Huogaro P, Bittolo BG, et al. Determination of high-density lipoproteins: screening methods compared. *Clin Chem.* 1979;25:939–942.
12. Lackland DT, Orchard TJ, Keil JE, et al. Are race differences in the prevalence of hypertension explained by body mass and fat distribution? *Int J Epidemiol.* 1992;21:236–245.
13. *SAS: Statistics, Version 5.* Cary, NC: SAS Institute; 1985.
14. National High Blood Pressure Education Program. The 1988 Report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure. *Arch Intern Med.* 1988;148:1023–1039.
15. Linn S, Fulwood R, Rifkind B, et al. High-density lipoprotein cholesterol levels among US adults by selected demographic and socioeconomic variables: the Second National Health and Nutrition Examination Survey 1976–1980. *Am J Epidemiol.* 1989;129:281–294.
16. Sprafka JM, Norsted SW, Folsom AR, Burke GL, Luepker RV. Life-style factors do not explain racial differences in high-density lipoprotein cholesterol: the Minnesota Heart Survey. *Epidemiology.* 1992;3:156–163.
17. Terry RB, Wood PD, Haskell WL, Stefanik ML, Krauss RM. Regional adiposity patterns in relation to lipids, lipoprotein cholesterol, and lipoprotein subfraction mass in men. *J Clin Endocrinol Metab.* 1989;68:191–199.
18. Forman MR, Trowbridge FL, Gentry EM,

et al. Overweight adults in the United States: the Behavioral Risk Factor Surveys. *Am J Clin Nutr.* 1986;44:410–416.
19. Najjar MF, Rowland M. Anthropometric reference data and prevalence of overweight: United States, 1976–1980. *Vital Health Stat [11].* 1987;238. DHHS publication PHS 87–1688.
20. Hartz A, Grubb B, Wild R, et al. The association of waist-to-hip ratio and angiographically determined coronary artery disease. *Int J Obesity.* 1990;14:657–665.
21. Larrson B, Svardsudd K, Welin L, Wilhelmsen L, Bjorntorp P, Tibblin G. Abdominal adipose tissue distribution, obesity, and risk of cardiovascular disease and death: 13-year follow-up of participants in the study of men born in 1913. *Br Med J.* 1984;288:1401–1404.
22. Ostlund RE, Staten M, Kohrt WM, Schultz J, Malley M. The ratio of waist-to-hip circumference, plasma insulin level, and glucose tolerance as independent predictors of the HDL2 cholesterol level in older adults. *N Engl J Med.* 1990;322: 229–234.
23. Folsom AR, Burke GL, Ballew C, et al. Relation of body fatness and its distribution to cardiovascular risk factors in young Blacks and Whites, the role of insulin. *Am J Epidemiol.* 1989;130:911–924.
24. Williams PT, Wood PD, Krauss RM, et al. Does weight loss cause the exercise-induced increase in plasma high-density lipoproteins? *Atherosclerosis.* 1983;47: 173–185.
25. Coon PJ, Bleecker ER, Drinkwater DT, Meyers DA, Goldberg AP. Effects of body composition and exercise capacity on glucose tolerance, insulin, and lipoprotein lipids in healthy older men: a cross-sectional and longitudinal intervention study. *Metabolism.* 1989;38:1201–1209.
26. Freedman DS, Strogatz DS, Eaker E, Joesoef MR, DeStefano F. Differences between Black and White men in correlates of high-density lipoprotein cholesterol. *Am J Epidemiol.* 1990;132:656–669.
27. Wilson PWF, Savage DD, Castelli WP, et al. HDL cholesterol in a sample of Black adults: the Framingham Minority Study. *Metabolism.* 1983;32:328–332.
28. Winkleby MA, Fortmann SP, Barrett DC. Social class disparities in risk factors for disease: eight-year prevalence patterns by level of education. *Prev Med.* 1990; 19:1–12.

Cancer Screening among African-American Women: Their Use of Tests and Social Support

ABSTRACT

This study examined associations between social ties (as measured by the Social Network Index), instrumental and emotional support, and the use of three female cancer screening tests: mammography, cervical smear, and clinical breast examination. Data were taken from a household survey of 670 African-American women living in northern California in 1986. In multiple logistic regression models, Berkman's index was associated with increased use of mammography but not with the use of cervical smear or clinical breast examination. Instrumental and emotional support measures were not significant. These results suggest that social networks may have a role in early cancer detection. (*Am J Public Health.* 1994;84:101–105)

Soo Hyang Kang, DrPH, Joan R. Bloom, PhD, and Patrick S. Romano, MD, MPH

Introduction

One explanation of why some people seek preventive care, such as cancer screening, is that their social ties may provide the necessary support or encouragement. Previous studies have produced conflicting results on how social ties are related to the use of preventive health services. Social ties have been positively associated,[1–5] not associated,[6] or negatively associated with preventive health behavior or services[7,8] in different published reports. Comparisons of these studies are problematic because of small sample sizes and different definitions of social ties and outcome measures. Few studies have explored associations between social ties and the use of preventive health services among African Americans.[9,10] The present study examines social ties (as measured by a social network index) and instrumental and emotional support and their relationships with the use of cancer screening tests among African-American women.

Methods

Self-reported use of mammography, cervical cytology, and clinical breast examination was investigated in a two-stage random household sample of the African-American population of Oakland and San Francisco, Calif, in 1986.[7,11] First, 100 census blocks in which African Americans constituted 25% or more of the population (based on 1980 census data) were randomly selected from each of the two cities. Second, households were randomly selected within these blocks. The Kish[12] procedure was used to randomly select one respondent 20 years of age or older from each African-American household, and face-to-face interviews were conducted. The response rates for Oakland and San Francisco were 67.6% and 69.1%, respectively. Of the 1137 respondents, 467 were men, who were excluded from the current study. Consistent with census data, 59% of the respondents were women (n = 670).

Soo Hyang Kang and Joan R. Bloom are with the School of Public Health, University of California, Berkeley. At the time of the study, Patrick S. Romano was with the Institute for Health Policy Studies, University of California, San Francisco. He is now with the School of Medicine, University of California, Davis.

Requests for reprints should be sent to Soo Hyang Kang, DrPH, 5378 Bryant Ave, Oakland, CA 94618.

This paper was accepted September 27, 1993.

TABLE 1—Distribution of Sample's Sociodemographic Characteristics (n = 670)

	%
Age, y	
Less than 35	37.6
35 to 55	34.2
55 and older	28.2
Health status	
Excellent	16.7
Good	45.7
Fair	30.3
Poor	7.3
Education	
Less than high school	24.1
High school graduate or equivalent	36.0
1–3 years of college	31.3
4 or more years of college	8.6
Health insurance coverage	
Health maintenance organization	37.7
Medi-Cal	19.7
Other insurance	33.6
None	9.0
Has a source of primary care	
Yes	93.9
No	6.1

Note. Totals may vary because of missing numbers.

TABLE 2—Number of Subjects Who Had a Routine Mammography, Cervical Smear, and Clinical Breast Examination, by Each Value of the Social Network Index

	Social Network Index, No. (%)					
Cancer Screening Test	Low	Medium Low	Medium High	High	Total	χ^2
Routine mammogram[a]						
Yes	11 (18)	54 (25)	13 (30)	35 (39)	113 (27)	0.029*
No	49 (82)	161 (75)	30 (70)	55 (61)	...	
Routine cervical smear[b]						
Yes	111 (91)	313 (91)	66 (94)	121 (94)	611 (91)	0.683
No	11 (9)	30 (9)	4 (6)	8 (6)	...	
Routine clinical breast examination[b]						
Yes	103 (84)	282 (83)	61 (88)	116 (90)	562 (84)	0.209
No	19 (16)	59 (17)	8 (12)	13 (10)	...	

[a] n = 418.
[b] n = 670.
* $P < .05$ (two-tailed test).

The survey instrument included questions on previous use of six cancer screening tests, social ties, instrumental and emotional support, health status, a source of primary care, insurance, and sociodemographic characteristics. Since the purpose of this study was to examine the use of female cancer screening tests as a preventive measure, only respondents who had a screening test as part of a routine examination were classified as having received the test; other respondents were classified as not having received the test. The three dependent variables in the study were whether or not the subject received (1) mammography, (2) cervical smear, and (3) clinical breast examination.

Social ties were measured by a slightly modified version of Berkman and Syme's Social Network Index. This index includes four different measures of social ties: marital status, number of relatives and friends, church participation, and participation in other organizations.[3] Possible scores range from a low of one (representing those with the fewest social ties) to a maximum of four (representing those with the most social ties). Emotional and instrumental support measures were adapted from Seeman and Syme.[13]

Respondents' self-reported health status, a source of primary care, health insurance coverage, age (along with a quadratic term to adjust for nonlinearity), and educational status were included as covariates because they have been previously related to the use of preventive health services.[7,14,15]

Multiple logistic regression models were estimated to determine whether social ties influenced the use of cancer screening tests after potential confounders were controlled. All female respondents were included in the model for routine cervical smear and routine clinical breast examination (n = 670). For routine mammography, only female respondents 35 years of age or older (n = 418) were included, since mammography is generally not recommended for women less than 35 years old.[16,17]

Results

Since the samples from the two cities were almost equal in size and showed similar demographic characteristics, they were combined for purposes of these analyses (Table 1). The results of the bivariate analyses between the Social Network Index and the use of cancer screening tests are presented in Table 2.

In a multiple logistic regression analysis for routine mammography (Table 3), the Social Network Index remained a significant explanatory variable after health status, age, age squared, education, having health insurance, and having a primary care source were controlled. The adjusted odds ratio (OR) was 1.3 for each additional level of the index (95% confidence interval [CI] = 1.03, 1.70), suggesting that women with larger social networks were more likely to have had a routine mammogram than those with smaller networks.

The Social Network Index was not a statistically significant predictor of the use of cervical smear (adjusted OR = 1.2, 95% CI = 0.9, 1.7) or clinical breast examination (adjusted OR = 1.1, 95% CI = 0.9, 1.5) in these analyses. The instrumental and emotional support measures were tested in separate logistic models with the same control variables. Neither of these support variables was significantly associated with the use of mammography, cervical smear, or clinical breast examination.

Discussion

Women with more social ties were more likely to have had a routine mammogram than were those with fewer social ties, even after health status, age, education, type of health insurance, and having a primary care source were controlled. The lack of any associations between the other two tests and the selected predictors may have been due to insufficient statistical power, since almost all women in this study had a cervical smear or a clinical breast examination. Another explanation may be the location of the screening. A mammogram generally requires a referral to a different facility, while both cervical smears and clinical breast examinations are performed during routine physical examinations. Social networks may be important not only in encouraging an initial

TABLE 3—Estimated Logistic Regression Coefficients and Odds Ratios Comparing Each Level of the Social Network Index ($n = 416$) for the Use of Routine Mammography

Predictor	Coefficient	SE	Adjusted Odds Ratio	95% Confidence Interval
Intercept	−12.85**	2.61	...	...
Health status (good and excellent = 1)	−0.09	0.27	0.91	0.54, 1.55
Social Network Index	0.28*	0.13	1.32*	1.03, 1.70
Age	0.36**	0.09	...	...
Education				
Less than high school	−0.17	0.31	0.84	0.46, 1.54
High school graduate or equivalent	Referent	...	...	...
1–3 years of college	−0.24	0.26	0.79	0.39, 1.59
4 or more years of college	0.27	0.50	1.31	0.49, 3.47
Health insurance coverage				
Health maintenance organization	Referent	...	...	...
Medi-Cal	−0.53	0.30	0.59	0.28, 1.26
Other insurance	−0.72*	0.29	0.49	0.28, 0.86
None	−0.64	0.50	0.53	0.20, 1.40
Source of primary care (having a source of primary care = 1)	0.44	0.80	1.55	0.33, 7.30
Age squared[b]	−0.0027**	0.006	...	...

Note. $n = 369$; model $\chi^2 = 56.90$ ($df = 11$); $-2\log L = 401.32$. Twenty-five observations were deleted because of missing values.
[a]Adjusted odds ratio associated with a one-unit change in the level of the Social Network Index.
[b]Because the linear and quadratic terms for age were both significant in this model, the odds ratio associated with 5-year age difference varies with age (e.g., the odds ratio comparing a 40-year-old woman with a 35-year-old woman would be 2.2, whereas the odds ratio comparing a 60-year-old woman with a 55-year-old woman would be 1.3).
*$P < .05$ (two-tailed test); **$P < .01$ (two-tailed test).

visit to the doctor's office but in following through with scheduling and receiving the test and in gaining acceptance for the doctor's recommendation.[14]

No associations were found between the instrumental and emotional support measures and the use of cancer screening tests. These measures were based on a small number of items and may not have captured the qualitative features of supportive social interactions. Alternatively, functional aspects of social support may be less important than the structural features measured by the Social Network Index.

Whether unmeasured factors, such as prior illness or personality characteristics, influence both social networks and use of cancer screening tests is not clear in these data. Chronic illness may lead to social isolation as people become unable to maintain personal contacts; it may also increase the use of diagnostic medical procedures. Our study findings are limited to the use of cancer screening tests among African Americans; additional studies are needed to determine whether these results apply to other ethnic and racial groups. Despite these limitations, this study suggests that African-American women with smaller social networks are less likely to receive a mammogram than are African-American women with larger networks. Interventions that use social ties may, therefore, have a role in promoting early cancer detection. □

References

1. Norman RMG, Tudiver F. Predictors of breast self-examination among family practice patients. *J Fam Pract.* 1986;22:149–153.
2. Langlie JK. Social networks, health beliefs, and preventive health behavior. *J Healt. Soc Behav.* 1977;18:244–260.
3. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine year follow-up study of Alameda County residents. *Am J Epidemiol.* 1979;109:186–204.
4. Zapka JG, Stoddard AM, Costanza ME, Greene HL. Breast cancer screening by mammography: utilization and associated factors. *Am J Public Health.* 1989;79:1499–1502.
5. Harlan LC, Bernstein AB, Kessler LG. Cervical cancer screening: who is not screened and why? *Am J Public Health.* 1991;81:885–890.
6. Rutledge DN. Factors related to women's practice of breast self-examination. *Nurs Res.* 1987;36:117–121.
7. Romano PS, Bloom JR, Syme SL. Smoking, social support, and hassles in an urban African-American community. *Am J Public Health.* 1991;81:1415–1422.
8. Mermelstein R, Cohen S, Lichtenstein E, Baer JS, Kamarck T. Social support and smoking cessation and maintenance. *J Consult Clin Psychol.* 1986;54:447–453.
9. Broadhead WE, Gehlbach SH, Degruy FV, Kaplan BH. Functional versus structural social support and health care utilization in a family medicine outpatient practice. *Med Care.* 1989;27:221–233.
10. Neighbors HW, Jackson JS. The use of informal and formal help: four patterns of illness behaviors in the Black community. *Am J Community Psychol.* 1984;12:629–644.
11. Bloom JR, Hayes WA, Saunders F, Flatt S. Cancer awareness and secondary prevention practices in Black Americans: implications for intervention. *Fam Community Health.* 1987;10(3):19–30.
12. Kish L. Procedure for objective respondent selection within the household. *J Am Stat Assoc.* 1949;44:380–387.
13. Seeman TE, Syme SL. Social networks and coronary artery disease: a comparison of the structure and function of social relations as predictors of disease. *Psychosom Med.* 1987;49:341–354.
14. Bloom JR, Grazier K, Hodge F, Hayes WA. Factors affecting the use of screening mammography among African American women. *Cancer Epidemiol Biomarkers Prev.* 1991;1:75–82.
15. Gluck ME, Wagner JL, Duffy BM. *The Use of Preventive Services by the Elderly.* Washington, DC: Office of Technology Assessment; 1989:77–101.
16. American Cancer Society. The American Cancer Society guidelines for the cancer-related checkup. *CA Cancer J Clin.* 1992; 42:44–45.
17. US Preventive Services Task Force. *Guide to Clinical Preventive Services.* Baltimore, Md: Williams & Wilkins; 1989:44.

Promoting Mammography Use through Progressive Interventions: Is It Effective?

Eunice S. King, PhD, RN, Barbara K. Rimer, DrPH, Janet Seay, MPH, Andrew Balshem, BA, and Paul F. Engstrom, MD

ABSTRACT

This study evaluated interventions implemented with women in a health maintenance organization who had not utilized their annual free mammogram referral 45 days (step 2) and 95 days (step 3) after its mailing. The step 2 evaluation compared mammography utilization for women randomly assigned to receive a brief reminder letter or no reminder: 42% of the reminder group vs 28% of the control group obtained mammograms. The step 3 evaluation compared utilization for women randomly assigned to receive a second reminder, a letter suggesting a preventive office visit, or telephone counseling: 29% of those who received telephone counseling, 14% of those who received a letter, and 12% of those who received a second reminder obtained mammograms. (*Am J Public Health.* 1994;84:104–106)

Introduction

Although regular mammograms can reduce mortality from breast cancer by about 34% in women aged 50 and older,[1,2] most American women are not getting them.[3–5] The purpose of this study was to evaluate interventions developed for an individual practice association–model health maintenance organization's (HMO's) annual breast cancer screening program directed at more than 50 000 women aged 50 through 74 years.[6] Interventions were implemented in three steps, beginning with a breast cancer information packet (step 1) containing a free mammogram referral (Table 1). A central office tracked mammography use through radiology reports and identified women eligible for subsequent interventions.

This study assessed whether a brief reminder (step 2) was more effective than no reminder and compared the relative impact of three different interventions—telephone counseling, a second reminder letter, and a letter suggesting a preventive office visit—in motivating still nonadherent women to get mammograms (step 3).

Methods

Procedures

For the step 2 and step 3 intervention evaluations, two mutually exclusive samples were used. So that both interventions could be completed within the same time period, thereby minimizing effects of secular events, women for the step 3 evaluation were selected first because of step 3's longer follow-up period (185 days vs 95 for step 2).

Step 2 evaluation. At the time their annual packet was mailed (step 1), 2127 women were randomized to a no-reminder or reminder group. Women in the latter group who did not obtain a mammogram within 45 days of the packet mailing were mailed a brief reminder letter. Ninety-five days after the packet mailing, a brief telephone survey was conducted to assess the women's mammography status at days 45 and 95. Of the original 2127 women, 745 were eligible for evaluation; of these, 21% could not be contacted during the allotted time interval, 19% had already had their mammograms, and 1% refused.

Step 3 evaluation. An annual packet (step 1) was mailed to 2123 women who constituted the initial sample and a step 2 reminder was sent to women without a mammogram report. Ninety days after the packet mailing, we attempted to survey 1265 women without a mammogram report to verify their mammography status. Of these, 25% could not be contacted during the 1-week eligibility evaluation window, 25% had already had mammograms, and 2% refused. Thus, 598 eligible women were randomized to one of three interventions: (1) a second reminder letter (n = 198); (2) a preventive office visit letter, urging women to have a checkup (n = 198); and (3) telephone counseling (n = 202, with 173 completed calls).

A telephone counseling intervention, based on the health belief model,[7] the transtheoretical model of behavior change,[8] and the conflict model of decision making,[9] was developed to help women overcome their personal barriers to mammography. The counseling content was guided by scripted responses to 26 potential barriers.[6,10–12]

Study Measures

The step 2 and step 3 telephone surveys, designed to obtain demographic data and the date and location of the woman's most recent mammogram, were conducted by professional interviewers in less than 5 minutes. The outcome variable was whether or not women obtained mammo-

Eunice S. King, Janet Seay, Andrew Balshem, and Paul F. Engstrom are with the Fox Chase Cancer Center, Cheltenham, Pa. Barbara K. Rimer is with the Duke Comprehensive Cancer Center, Durham, NC.

Requests for reprints should be sent to Eunice S. King, PhD, RN, Fox Chase Cancer Center, 510 Township Line Rd, Cheltenham, PA 19012.

This paper was accepted May 19, 1993.

Editor's Note. See related editorials by Shapiro (p 10) and Zapka (p 12) in this issue.

TABLE 1—Relationship between Annual Screening Program and Evaluation Procedures for Step 2 and Step 3 Interventions

	Annual Program	Step 2 Evaluation	Step 3 Evaluation
Step 1			
Day 1	Mail annual packet, including health education materials and referral for a free mammogram.	2127 packets mailed (4/29/91–7/1/91). Women randomized to step 2 reminder vs no reminder.	2123 packets mailed (3/11/91–4/23/91).
	If no mammogram, then proceed to step 2.	No mammogram report for 1524 women.	No mammogram report for 1710 women.
Step 2			
Day 45	Mail reminder letter.	783 mailed reminder, 761 no reminder.	1710 women mailed reminder.
Day 90	If no mammogram, then proceed to step 3.		Telephone survey to assess mammography status of 1285 women without reports.
Step 3			
Day 95	Random assignment: Second reminder letter Preventive visit letter Telephone counseling	Telephone survey to assess mammography status at day 45 (n = 1446). 745 eligible for step 2 intervention and evaluation: 381 mailed reminder, 364 no reminder.	686 eligible women randomized: Second reminder (n = 198) Preventive visit letter (n = 196) Telephone counseling (n = 202)

TABLE 2—Demographic Characteristics of Women in the Step 2 and Step 3 Evaluations

	Step 2		Step 3	
	No.	%	No.	%
Marital status				
Married	504	68	415	70
Not married	234	32	175	30
Employment				
Employed	534	73	395	67
Not employed	200	27	194	33
Education				
11th grade or less	126	17	104	18
High school	382	52	318	54
Beyond high school	222	30	164	28
Race				
White	641	88	491	84
Non-White	86	12	91	16
Age				
50–54 y	311	42	218	37
55–59 y	221	30	187	32
60–64 y	144	19	134	23
65–75 y	69	9	54	9

TABLE 3—Variables Related to Use of Mammogram Referral*

	Obtained Mammogram				
	Yes		No		
	No.	%	No.	%	P
Step 2 evaluation					
Study group					
No reminder	100	28	264	73	
Reminder	159	42	222	58	<.001
Marital status					
Married	193	38	311	62	
Not married	63	27	171	73	.003
Education					
Less than high school	33	26	93	74	
High school	129	34	253	66	
More than high school	69	42	129	58	.010
Step 3 evaluation					
Telephone counseling	57	28	143	72	
Letter	28	14	170	86	
Second reminder	23	12	173	88	<.001

*Based on chi-square analysis test of homogeneity.

grams following the interventions (between 46 and 95 days after the packet mailing for women in the step 2 evaluation and within 90 days following the intervention for women in step 3). Only women whose self-reports of mammography were verifiable by medical records were considered adherers.

Data Analysis

For both the step 2 and step 3 evaluations, chi-square tests of significance were used to assess relationships between study group and adherence and demographic characteristics and adherence. In the step 2 analysis, possible confounding by marital status and education was controlled by including these variables along with study group in a logistic regression analysis. In the step 3 analysis only study group was included, because no demographic characteristic was related to adherence.

Results

Most women in the step 2 and step 3 evaluations were married, employed, between 55 and 75 years of age, and had at least a high school education; 12% to 16% were African Americans (Table 2). There were no significant differences in the demographic characteristics of the study groups.

The step 2 evaluation found that 42% of the reminder group and 28% of the no-reminder group ($P < .001$) had obtained mammograms within the 46- to 95-day period (Table 3). As Table 4 shows, women who received the step 2 reminder were more likely to have had a mammogram than those who did not (odds ratio [OR] = 1.9). Women with more than a high school education were more likely than those with less education to have had a mammogram (OR = 2.0), and married women were more likely than unmarried women to have had one (OR = 1.7). There were no significant interactions.

In the step 3 evaluation, 29% of the women in the telephone counseling group obtained subsequent mammograms, compared with only 12% of those in the second reminder group and 14% in the preventive office visit letter group ($P < .001$). The odds of having had a mammogram were three times as high for women in the telephone counseling group as for women in the second reminder group, whose odds were similar to those of women in the preventive visit letter group (see Table 4).

Discussion

In this study, a simple, inexpensive letter ($0.91 per success) resulted in significant improvement in mammography use 46 to 95 days after the annual packet mailing. For women who were still nonadherent, telephone counseling, compared with the letters, nearly doubled the odds of getting a mammogram. The cost per success for telephone counseling ($4.92), although higher than for the second reminder ($2.73) or preventive letter

TABLE 4—Logistic Regression Analyses: Variables Associated with Having a Mammogram

	Odds Ratio	95% Confidence Interval	P
Step 2 evaluation			
Reminder vs no reminder	1.9	1.4, 2.6	.001
Married vs not married	1.7	1.2, 2.4	.003
High school graduate vs did not complete high school	1.4	0.9, 2.1	.2
More than high school vs did not complete high school	2.0	1.2, 3.2	.006
Step 3 evaluation			
Phone counseling vs second reminder	3.0	1.8, 5.1	< .001
Preventive visit letter vs second reminder	1.2	0.7, 2.2	.48

($3.68), is still reasonably inexpensive. Telephone counseling conducted by volunteers, rather than professionals, may reduce the cost of this intervention and should be evaluated in future studies.

The generalizability of this study's results is limited by the fact that the study was conducted within an HMO population for whom the cost barrier was removed by the annual free mammogram referral. In addition, because there was no control group for the step 3 evaluation, we cannot separate the effect of the telephone survey to assess status from the effect of the interventions. However, it is likely that the potential impact of the telephone survey would be distributed equally across the three groups within the step 3 evaluation sample. In future studies other variables, such as women's underlying knowledge, attitudes, and current preventive health practices, should be assessed and the interventions should be tested in other (non-HMO) settings. □

Acknowledgments

This study was supported in part by National Institutes of Health grant S R01-CA45834-03.

Preliminary results of this study were presented by Dr Barbara Rimer at the 119th Annual Meeting of the American Public Health Association, Atlanta, Ga, November 13, 1991.

We would like to thank Susan Montgomery and Cheryl Baum, telephone counselors for this project; Robert McLaughlin and Linda Crossley from the US Healthcare Check office; Susan Montgomery and Terry Torres for word processing; Mary Daly, MD, PhD, Caryn Lerman, PhD, and Bruce Trock, PhD, for their review and comments; and George Bonney, PhD, for his statistical consultation.

References

1. Elwood JM, Cox B, Richardson AK. The effectiveness of breast cancer screening by mammography in younger women. *Online J Curr Clin Trials.* 1993;February 25(Doc no. 32).
2. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Selections, follow-up, and analysis in the Health Insurance Plan study: a randomized trial with breast cancer screening. *Monogr Natl Cancer Inst.* 1985;67:65–74.
3. Dawson DA, Thompson GB. Breast cancer risk factors and screening: United States, 1987. *Vital Health Stat [10].* 1990; no. 172:1–33. DHHS publication PHS 90-1500.
4. National Cancer Institute Breast Cancer Screening Consortium. Screening mammography: a missed clinical opportunity? Results of the NCI Breast Cancer Screening Consortium and National Health Interview Survey studies. *JAMA.* 1990;264:54–58.
5. Rakowski W, Rimer BK, Bryant SA. Integrating behavior and intention for the study of mammography: data from the 1990 supplement to the National Health Interview Survey. *Public Health Rep.* 1993;108: 605–624.
6. Rimer B, Ross E, Balshem A, Engstrom PF. The effect of a comprehensive breast screening program on self-reported mammography by primary care physicians and women in a health maintenance organization. *J Am Board Fam Pract.* 1993;108: 605–624.
7. Rosenstock IM. The health belief model: explaining health behavior through expectancies. In: Glanz K, Lewis FM, Rimer BK, eds. *Health Behavior and Health Education.* San Francisco, Calif: Jossey-Bass Publishers; 1990:39–62.
8. Prochaska J, DiClemente C. Stages and processes of self-change of smoking toward an integrative model of change. *J Consult Clin Psychol.* 1983;51:390–395.
9. Janis IL. *Counseling on Personal Decisions.* New Haven, Conn: Yale University Press; 1982.
10. Zapka JG, Stoddard A, Costanza ME, Greene HL. Breast cancer screening by mammography: utilization and associated factors. *Am J Public Health.* 1989;79: 1499–1502.
11. Rimer BK, Kasper-Keintz M, Kessler HB, Engstrom PF, Rosan JR. Why women resist screening mammography: patient-related barriers. *Radiology.* 1989;172:243–246.
12. Rimer BK, Davis SW, Engstrom PF, Myers R, Rosan JR. Some reasons for compliance and noncompliance in a health maintenance organization breast cancer screening program. *J Compliance Health Care.* 1988;3:103–114.

The Use of Mammography Vans by Low-Income Women: The Accuracy of Self-Reports

Susan Etzi, MD, MPH, Dorothy S. Lane, MD, MPH, and Roger Grimson, PhD

ABSTRACT

The objective of this study was to determine the accuracy of self-reports of mammography use by low-income women. Mammography van records were used to verify self-reports of mammography use in the past year by women aged 50 through 75 years who had visited five community health centers (n = 237). Van records verified mammography use for 99% of these women (82% within the previous year and 98% within the past 2 years). Forty percent of those with van records who reported both the month and year of the mammogram were accurate. Inaccurately reported dates were more frequently after (74%) rather than before (26%) the actual date. These findings indicate that self-reports of mammography use by low-income women are generally reliable. (*Am J Public Health.* 1994;84:107–109)

Introduction

National objectives and initiatives specifically target low-income women for interventions designed to increase mammography use because there is evidence of particularly low utilization rates among the disadvantaged.[1–3] Methods of measuring the effectiveness of interventions designed to increase mammography use are needed. There are reports of a high validity of women's self-reports in health maintenance organization (HMO)[4,5] and community[6] settings. We examined whether this is also true for a less educated, lower-income population that visits community health centers.

The Suffolk County Department of Health Services funds five health centers in the intervention area of our community-based breast cancer project.[7] Beginning in January 1988, mammography was provided at each of the health center sites, initially through a county contract with a private mammography van service and subsequently (beginning February 1990) through a mammography van purchased by the Department of Health Services.

Methods

A telephone survey was conducted in 1990 of a random sample of women aged 50 through 75 years from computer files of women who had visited one or more of the five health centers for any reason within the previous year. The survey methodology has been described in detail elsewhere.[8] Sixty-three percent of the original sample completed the survey. Lack of a telephone number was the most common reason for failure to contact the women: 19% of the sample did not have telephones or had unknown or incorrect telephone numbers. Of the women with apparently correct telephone numbers, who therefore could potentially be contacted, 75% completed the survey.

Of the 844 women interviewed, 806 were included in the analysis (38 were excluded because of a prior history of breast cancer or because they were not in the age range of 50 through 75 years). The socio-demographic characteristics of the 806 respondents have been described elsewhere; they were consistent with a disadvantaged population.[8] The survey instrument included questions about whether the respondent had ever had a mammogram, when and where she had her last mammogram, and who paid for it. The date of the last mammogram was recorded by month (or season if the month could not be recalled) and by year.

The validation study of women's self-reports was limited to the 237 women who indicated that they had had a mammogram in a van in the previous year. Matches between survey and van records were double-checked by hand to identify incorrect or missed matches due to spelling errors and for accuracy of match by patient's name, address, date of birth, and clinic site where patient received her usual medical care. The patient's estimated date of her last mammogram and the actual date recorded in the van records were compared. Reports of a mammogram in the previous year were verified by determining whether the actual date of the van mammogram was no more than 12 months prior to the survey month. Since the van files covered a period longer than 1 year, occasionally a patient had a record of more than one mammogram in the van files. The date on which the woman was surveyed by telephone was used to compute the 12-month period of interest, thereby determining which mammogram record was to be included in the study.

Results

Mammography Use

Of the 806 respondents, 631 (78%) reported ever having had a mammogram.

At the time of the study, all authors were with the Department of Preventive Medicine, School of Medicine, State University of New York at Stony Brook.

Requests for reprints should be sent to Dorothy S. Lane, MD, MPH, Department of Preventive Medicine, School of Medicine, State University of New York at Stony Brook, Stony Brook, NY 11794-8036.

This paper was accepted May 11, 1993.

Editor's Note. See related editorials by Shapiro (p 10) and Zapka (p 12) in this issue.

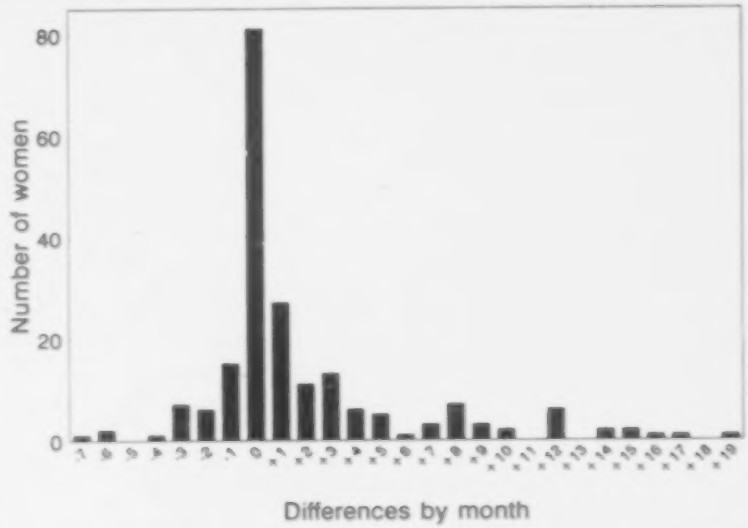
TABLE 1—Number of Women for Whom Documentation Supported the Accuracy of Self-Reported Mammography

Actual Date of Mammogram	No. (n = 237)	%
Within past 12 mo	194	81.9
Within past 13–24 mo	38	16.0
Within past 25–36 mo	3	1.3
No record found	2	0.8

TABLE 2—Number of Women Who Recalled Mammography Date Accurately, by Time since Mammogram

Time Since Mammogram	n	Estimated Date vs Real Date			Mean Difference[a]	Median Difference[a]
		No. Earlier	No. Same	No. Later		
≤3 mo	51	7	35	9	−2.24	0
>3–6 mo	54	12	24	18	1.15	0
>6–12 mo	71	13	21	37	45.92	30
>12 mo	28	0	0	28	297.86	288.5
Total	204	32	80	92		

[a]Difference = estimated date minus real date in months.



Note. 0 = agreement by month and year (accurate); − = estimated date before actual date; + = estimated date after actual date.

FIGURE 1—Differences between actual and estimated months of mammogram (n = 204).

Three hundred seventy-eight (47%) reported having had a mammogram in the previous year, and the majority of these (237, or 63%) gave a van as the location of the mammogram. Twenty percent said they had had the mammogram at a radiologist's or doctor's office; 12% at a clinic, health center, or HMO; and 6% at a hospital.

Accuracy of Reporting

Van records were obtained for 99% of the women who reported having had a mammogram in a van in the previous year, and 82% of these women had documentation that the mammogram was actually performed in that year (Table 1). An additional 16% had documentation of a van mammogram within the past 13 through 24 months, and 1% had documentation of a van mammogram within 25 through 36 months. No documentation was found for 2 women (0.8%).

Comparisons between self-reports and van records for month and year of the mammogram showed that 81 (40%) were in agreement, 124 (61%) were within 1 month, 161 (79%) were within 3 months, and 197 (97%) were within 1 year. For 7 (3%) of the women, the actual date differed from the self-reported date by more than a year. For the 123 women who were inaccurate in their recall, the self-reported date was more frequently after the actual date of mammography (for 91 women, or 74%) than prior to the actual date (for 32 women, or 26%). The difference between the actual and estimated dates of mammography in number of months is displayed in Figure 1.

Accuracy of recall diminished with time since the last mammogram,* as detailed in Table 2. The majority of women who had had a mammogram within the 3 months prior to being surveyed were accurate in their recall, whereas the majority of women who had had their mammograms from 6 to 12 months previously believed that they had had the mammogram later than the actual date. This was also true for all of the women who had actually had the mammogram more than a year prior to their date of recall.

Discussion

The results of our study indicate that the self-reports of low-income women visiting community health centers are a reliable source of data on mammography use in the past year. Van records verified mammography use for 99% of the women studied (82% within the prior year and 98% within the past 2 years). For 40% of the women who reported both the month and year of the mammogram, the dates corresponded to van records. The remaining women, who were inaccurate in recalling the date, were more likely to believe that they had had the mammogram more recently than they actually did (74%) rather than earlier. This memory effect is known as forward telescoping.[10,11] Accuracy in estimating the date of the mammogram declined with time,* and more than 6 months after having the mammogram most women (66%) believed that they had had it later than the actual date. More than 12 months after the mammogram, all women believed that they had had it later than the actual date. If the women were to rely on their own recall in scheduling subsequent mammograms, the time between mammograms would extend beyond the recommended interval. This suggests the desirability of using health center–initiated patient reminders to prompt timely scheduling of annual mammography.

A 94% accuracy rate for self-reports of mammography use in the past year

among women enrolled in an independent practice association HMO was reported by King et al.[4] Our project found that 88% of the self-reports of mammography use in the past year among women enrolled in a staff model HMO were confirmed by medical records.[5] This proportion is close to the 82% rate we found for the self-reports of women attending health centers, despite statistically significantly higher income and education in the HMO group than in the health center group. The women in the HMO also tended to underestimate rather than overestimate the time since their last mammogram.[5]

To explore the effectiveness of interventions designed to accomplish the Year 2000 objectives, it is important to be able to measure changing mammography screening practices among low-income women. The results of our validation study indicate that surveys of such women provide reliable data on mammography use in the previous year. □

Acknowledgments

This research was supported by grant 1R01 CA 45034-01 from the National Cancer Institute.

The authors thank Dr Andre Varma for technical assistance in linking the computer files of survey and van record data, Leijin Zhao for assistance in computer data analysis of the 1990 survey data, and Dr Lee Caplan for additional technical assistance. Dr David Kovenock of Northeast Research (Orono, Maine) provided assistance in revising the telephone surveys, which were conducted by contract with Northeast Research. We are indebted to Mary Solomon from the Women's Outreach Network for providing access to the Network's van records; to Dr Clare Bradley of the Suffolk County Department of Health Services for providing access to the county van data and health center registrants; and to the community advisory boards of the five health centers for approving our survey of the health center population.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-5021.
2. National Cancer Institute Breast Cancer Screening Consortium. Screening mammography: a missed clinical opportunity? Results of the NCI Breast Cancer Screening Consortium and National Health Interview Survey studies. *JAMA.* 1990;264: 54–59.
3. Centers for Disease Control. Use of mammography—United States, 1990. *MMWR.* 1990;39:621–630.
4. King ES, Rimer BK, Trock B, Balsham A, Engstrom P. How valid are mammography self-reports? *Am J Public Health.* 1990;80: 1386–1388.
5. Fulton-Keogh DL, Burg MA, Lane DS. Are self-reported dates of mammograms accurate? *Public Health Rev.* In press.
6. Degnan D, Harris R, Ranney J, Quade D, Earp JA, Gonzale J. Measuring the use of mammography: two methods compared. *Am J Public Health.* 1992;82:1386–1388.
7. Lane DS, Polednak AP, Burg MA. Measuring the impact of varied interventions on community-wide breast cancer screening. In: Anderson PN, Engstrom P, Mortonson LE, eds. *Advances in Cancer Control.* Vol. 6. New York, NY: Alan R. Liss Inc; 1989: 103–114.
8. Lane DS, Burg MA. Strategies to increase mammography utilization among community health center visitors: improving awareness, affordability and accessibility. *Med Care.* 1993;31:175–181.
9. Baddeley AD. *The Psychology of Memory.* New York, NY: Basic Books Inc; 1976:3–15,58–65,93–99.
10. Neter J, Waksberg J. A study of response errors in expenditure data from household interviews. *J Am Stat Assoc.* 1964;59:17–55.
11. Bremer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S, eds. *Measurement Errors in Surveys.* New York, NY: John Wiley & Sons Inc; 1991:134–137, 175.

WFPHA International Congress on Health, Economics, and Development

The World Federation of Public Health Associations (WFPHA) will hold its Eighth International Congress, "Health, Economics, and Development: Working Together for Change," December 5 through 9, 1994, in Bali, Indonesia. Join health practitioners, policymakers, administrators, development workers, economists, researchers, and many others from governments, international organizations, and private groups in this exciting and ambitious event. Examine the linkages among health, economics, and development. Develop an action agenda to integrate health concerns with economic and development policies.

The WFPHA is a nongovernmental organization composed of national public health associations from 41 countries around the world. WFPHA congresses are held every 3 years and are cosponsored by the World Health Organization and the United Nations Children's Fund. This congress will be hosted by the Indonesian Public Health Association. Participation is open to all. The official language of the congress is English.

For further information and abstract forms, contact WFPHA Secretariat, c/o American Public Health Association, 1015 15th St NW, Suite 300, Washington, DC 20005 USA; fax (202) 789-5681. The deadline for abstract submission is *May 15, 1994.*

Screening for Childhood Lead Poisoning: A Cost-Minimization Analysis

ABSTRACT

Decision analysis was used to compare the costs of three screening strategies for childhood lead poisoning: (1) venipuncture; (2) capillary sample with venipuncture confirmation if the blood lead level is elevated; (3) stratification by risk, with venipuncture for high-risk children and capillary sample for low-risk children. Under baseline conditions, the cost of screening by the venipuncture, stratification, and capillary strategies is $22, $25, and $27, respectively. Venipuncture remains the least expensive strategy unless the cost of venipuncture is more than three times that of capillary sampling. The annual cost of a national lead screening program that uses a single venipuncture sample would be $352 million. Initial screening with a capillary sample would cost $432 million, 23% more than venipuncture. (*Am J Public Health.* 1994;84:110–112)

Deborah E. Glotzer, MD, MPH, Howard Bauchner, MD, Kenneth A. Freedberg, MD, MSc, and Sean Palfrey, MD

Introduction

Lead poisoning is one of the most common and preventable illnesses of childhood. It is estimated that one of every six United States children has an elevated blood lead level (≥0.72 μmol/L [15 μg/dL]), which is associated with adverse health effects.[1] The 1991 Centers for Disease Control (CDC) guidelines for lead screening differ significantly from earlier versions.[2] The CDC now recommends universal blood lead screening for all US children under age 6 years, except for children in communities that can demonstrate that they do not have a childhood lead poisoning problem. A blood lead level is now recommended as the primary screening test, rather than the erythrocyte protoporphyrin level, because erythrocyte protoporphyrin screening is not sufficiently sensitive to detect most children with blood lead levels above 1.21 μmol/L (25 μg/dL).[3–5]

The guidelines advise that a level of 0.72 μmol/L (15 μg/dL) or greater from a capillary sample should be confirmed by a venous sample because capillary samples may be falsely positive. Between 22% and 29% of capillary samples were found to be falsely elevated when the blood lead level of concern was 1.93 μmol/L (40 μg/dL),[6] so the false-positive rate is likely to be at least as high if the threshold is lower. However, obtaining specimens by venipuncture is more difficult than capillary sampling, and initial screening by venipuncture is not widely used.[7] The objective of this study was to compare the costs of universal blood lead screening by different strategies.

Methods

We used decision analysis[8,9] to compare the costs of the following three strategies: (1) initial screening with a venipuncture specimen; (2) initial screening with a capillary specimen, with venipuncture confirmation if the blood lead level was ≥0.72 μmol/L (15 μg/dL); and (3) stratification according to risk level as determined by questionnaire and then screening high-risk children with a venipuncture specimen and low-risk children with a capillary specimen, with venipuncture confirmation for blood lead levels of ≥0.72 μmol/L (15 μg/dL).

Baseline estimates of variables in the model were obtained from the medical literature. Costs were obtained from the 1990 rate book of The Children's Hospital, Boston, Mass.[10] This rate book lists the same cost ($3) for phlebotomy by both venipuncture and capillary sampling. However, in the baseline analysis we estimated that the cost of phlebotomy by venipuncture was twice as expensive as capillary sampling to more closely simulate the national experience. Costs rather than charges were used for this analysis, because costs more accurately reflect true resource use. SMLTREE software[11] was used to construct and evaluate the decision tree. Sensitivity analysis was used to explore the effect of changing variables over a range of clinically reasonable values (Table 1).

Results

Given baseline prevalences and costs, initial screening of all children with a venipuncture specimen cost $22 per child, and screening by the stratification and capillary strategies cost $25 and $27, respectively.

Sensitivity analysis (Table 2) demonstrated that the venipuncture strategy was the least expensive strategy over almost the entire range of variables indicated in

Deborah E. Glotzer, Howard Bauchner, and Sean Palfrey are with the Division of General Pediatrics, Department of Pediatrics, and Kenneth A. Freedberg is with the Clinical Economics Research Unit, Department of Medicine and Thorndike Memorial Laboratory, Boston City Hospital and Boston University School of Medicine, Boston, Mass.

Requests for reprints should be sent to Deborah E. Glotzer, MD, MPH, Department of Pediatrics, Boston City Hospital–Talbot 102, 818 Harrison Ave, Boston, MA 02118.

This paper was accepted May 27, 1993.

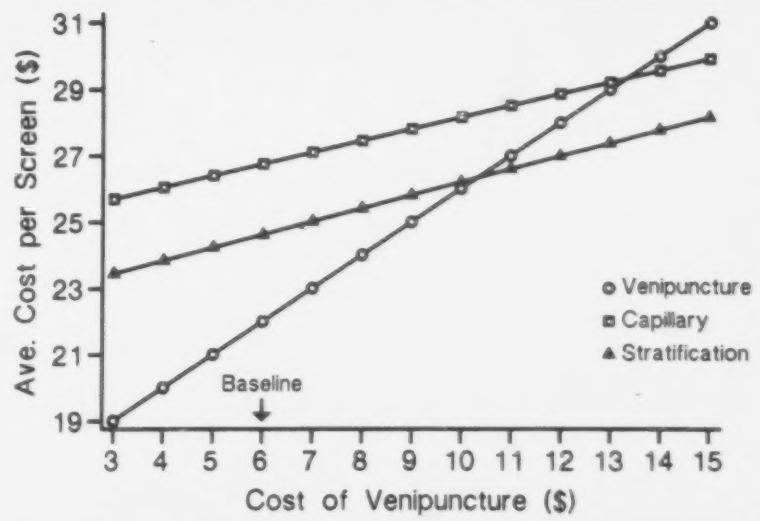
TABLE 1—Baseline Prevalence and Cost Estimates in the Model and Ranges Tested in Sensitivity Analysis

	Baseline (Range)	Reference
Prevalence of blood lead levels of ≥ 0.72 µmol/L (15 µg/dL) for all children	17% (1–50%)	1
Prevalence of blood lead levels of ≥ 0.72 µmol/L (15 µg/dL) for low-risk children	7.5% (0.5–10%)	1
Percentage of high-risk children	16% (1–50%)	1
False-positive rate	22% (5–50%)	6
Cost of phlebotomy for venipuncture specimen	$6 ($3–15)	10 and authors' estimate
Cost of phlebotomy for capillary specimen	$3 ($1–10)	10
Cost of blood lead analysis	$16 ($5–75)	10
Cost of erythrocyte protoporphyrin analysis	$0 ($0–15)	10

TABLE 2—Sensitivity Analysis of Selected Variables: Cost per Child According to Screening Strategy

	Value Tested	Cost of Screening Strategy, $		
		Venipuncture	Stratification	Capillary
Cost of blood lead analysis (baseline value = $16)	$5	11	11	12
	$35	41	48	52
	$75	81	97	107
Cost of phlebotomy for venipuncture specimen (baseline value = $6)	$3	19	23	26
	$9	25	26	28
	$15	31	28	30
Prevalence of blood lead levels of ≥ 0.72 µmol/L (15 µg/dL) (baseline value = 17%)	1%	22	24	24
	5%	22	24	25
	30%	22	25	29
Cost of phlebotomy for venipuncture specimen ($10) and prevalence of blood lead levels of ≥ 0.72 µmol/L (5%) (baseline values = $6 and 17%, respectively)		26	25	26
False-positive rate (baseline value = 22%)	5%	22	22	24
	35%	22	27	29
Cost of erythrocyte protoporphyrin analysis (baseline value = $0)	$5	27	31	34
	$10	32	37	40
	$15	37	43	47

Note. See Table 1 for baseline values for all variables.



Note. Venipuncture is the least expensive strategy unless the cost of venipuncture exceeds $10 per sample. The capillary strategy is more costly than the venipuncture strategy if the cost of venipuncture is less than $13 per sample. In the baseline analysis, the costs of venipuncture and capillary sampling were estimated at $6 and $3 per sample, respectively. Ave. = average.

FIGURE 1—The cost of lead screening plotted as a function of the cost of venipuncture.

Table 1. Because venipuncture is the final step in all the screening strategies, and because the cost of laboratory analysis is the same for capillary and venous samples, the venipuncture strategy is the least expensive strategy unless the cost of venipuncture phlebotomy is more than three times higher than the cost of capillary sampling (Figure 1).

As the prevalence of elevated blood lead levels decreased from 17% (baseline) to 5%, the cost of the venipuncture strategy was unchanged ($22). However, the cost of the stratification ($24) and capillary ($25) strategies decreased, reducing the added cost of these strategies compared with the venipuncture strategy. Under conditions of low prevalence (5%) and a high cost of venipuncture ($10), the stratification strategy ($25) became less expensive than either the venipuncture ($26) or capillary ($26) strategies.

If only 5% of the capillary samples are falsely elevated, the cost of screening by venipuncture would be the same as the cost of the stratification strategy ($22), but would remain less expensive than the capillary strategy ($24).

In our baseline analysis, we assumed that only blood lead analysis would be performed. However, erythrocyte protoporphyrin levels are often used as an adjunctive test for managing childhood lead poisoning. If erythrocyte protoporphyrin testing ($10) is included in the screening strategies, the venipuncture strategy ($32) would become even less expensive relative to the stratification ($37) and capillary ($40) strategies.

Discussion

This analysis demonstrates that a single venipuncture sample is the least

costly method of universal screening for childhood lead poisoning. The CDC recommends annual screening for children at low risk for lead poisoning and screening at 6-month intervals for high-risk children. Based on the US Public Health Service estimates,[1] 16 million blood lead screening tests on children under age 6 years will be required annually. According to our model, a lead screening program that uses a single venipuncture sample would cost $352 million per year. The annual costs of screening by the risk stratification and capillary strategies would be $400 million and $432 million, respectively, a 14% to 23% increase over the cost of screening with a single venipuncture sample.

Sensitivity analysis demonstrated that the venipuncture screening strategy is the least expensive approach over a wide range of prevalences, false-positive rates, and costs. For populations with a low prevalence of elevated blood lead levels, initial screening with capillary specimens may appear to be a more attractive strategy, at minimal additional average cost. However, because of the large numbers of screening samples required in areas of even moderate prevalence and the high rate of necessary rescreening because of false-positive results, the overall additional annual expenditure generated by this approach is considerable.

Issues of time, availability of skilled pediatric phlebotomists, convenience, and child/parent preference for the method of sampling were not considered in this analysis. Although initial screening by capillary sample may appear to be simpler than initial screening by venipuncture, that may not be true if a confirmatory sample by venipuncture is required.

This analysis indicates that methods for childhood lead screening differ substantially in cost. Better data are needed regarding the false-positive rate by capillary sampling now that the blood lead level requiring rescreening is 0.72 μmol/L (15 μg/dL). In addition, improved capillary sampling techniques that eliminate the need for venous confirmatory samples are needed. However, based on current information and existing techniques, a screening strategy that uses a single venipuncture sample can reduce the cost of screening for childhood lead poisoning. □

Acknowledgments

This research was supported in part by grants from the Health Resources and Services Administration, Bureau of Health professions: D28PE11122 (Residency Training in General Pediatrics) and D28PE51008 (Faculty Development in General Pediatrics).

This paper was presented in part at the 32nd Annual Meeting of the Ambulatory Pediatric Association, Baltimore, Md, May 5, 1992.

The authors wish to thank Ellen Coates for help with manuscript preparation.

References

1. Agency for Toxic Substances and Disease Registry. *The Nature and Extent of Lead Poisoning in Children in the United States: A Report to Congress.* Atlanta, Ga: US Dept of Health and Human Services; 1988.
2. Centers for Disease Control. *Preventing Lead Poisoning in Young Children: A Statement by the Centers for Disease Control.* Atlanta, Ga: US Dept of Health and Human Services; 1991.
3. Mahaffey KR, Annest JL. Association of erythrocyte protoporphyrin with blood lead level and iron status in the second national health and nutrition examination survey, 1976–1980. *Environ Res.* 1986;41:327–338.
4. Turk DS, Schonfeld OJ, Cullen M, Rainey P. Sensitivity of erythrocyte protoporphyrin as a screening test for lead poisoning. *N Engl J Med.* 1992;326:137–138.
5. DeBaun MR, Sox HC. Setting optimal erythrocyte protoporphyrin screening decision threshold for lead poisoning. *Pediatrics.* 1991;88:121–131.
6. Mitchell DG, Aldous KM, Ryan FJ. Mass screening for lead poisoning. *N Y State J Med.* 1974:1599–1603.
7. Edwards KS, Forsyth BWC. Lead screening at pediatric teaching programs. *Am J Dis Child.* 1989;143:1455–1457.
8. Pauker SG, Kassirer JP. Decision analysis. *N Engl J Med.* 1987;316:250–258.
9. Weinstein MC, Fineberg HV. *Clinical Decision Analysis.* Philadelphia, Pa: WB Saunders Co; 1980.
10. *1990 Rate Book.* Boston, Mass: Children's Hospital; 1990.
11. *SMLTREE, Version 2.9.* Roslyn, NY: Jim Hollenberg; 1989.

Who Participates in Cardiovascular Disease Risk Factor Screenings? Experience with a Religious Organization–Based Program

Barbara L. Wells, PhD, Charles C. Brown, PhD, John W. Horm, MSc, Richard A. Carleton, MD, and Thomas M. Lasater, PhD

ABSTRACT

Adult members who declined participation in cardiovascular disease risk factor screenings offered at religious organizations were randomly selected and asked to participate in screenings at their homes. Relationships between screening participation and sociodemographic, behavioral, and physiological measures were examined. Age, knowledge of cardiovascular disease risk factors, body mass index, current smoking status, previous report of elevated blood pressure, current diastolic blood pressure measurement, frequency of worship service attendance, and residential distance from the religious organization screening site were important predictors of screening response. Those with conspicuous risk factors appeared less likely to initially respond to religious organization site screening invitations. (*Am J Public Health.* 1994;84: 113–115)

Introduction

Cardiovascular diseases continue to be the leading cause of death in the United States.[1] Among their modifiable risk factors, elevated blood cholesterol and blood pressure are generally asymptomatic and present specific detection issues. Screening is essential for the identification and referral of undetected, untreated, or inadequately treated hypertensive or hypercholesterolemic people.[2,3]

One promising community setting for conducting screenings is religious organizations.[4] More than half the US population are reported to be members of an estimated 286 811 religious congregations,[5,6] and religious organization screening programs that target elevated blood pressure among Black[7–13] and White populations[14] have reported generally favorable, but not unanimous, results. Nevertheless, there have been those individuals who have chosen not to participate in such screenings.

Few studies have characterized screening nonparticipants for any of the cardiovascular disease risk factors. However, because the research that forms the context of this report was part of a multiple cardiovascular disease risk factor intervention trial testing the efficacy of religious organizations as health promotion sites, that trial presented the opportunity to examine the sociodemographic, behavioral, and physiological differences between participants and nonparticipants.

Methods

The intervention trial was conducted in 20 religious organizations randomly selected from the largest denominations in the state of Rhode Island.[15] The 10 Catholic, 5 Northern Baptist, and 5 Episcopal religious organizations are in 15 communities and have predominantly Caucasian memberships of between 250 and 5000 individuals. Approximately 10% of each organization's members aged 18 to 65 years were randomly selected for the initial screening. A letter of invitation from the pastor requested each selected member to schedule an appointment, at that person's convenience, for a free cardiovascular disease risk factor screening. At least five telephone contact attempts were made to those who did not call before a potential participant was deemed not reachable.

Ten religious organizations, proportionally representing the religious denomination distribution of participants, were selected to participate in the follow-up survey to assess nonparticipation. The nonparticipant survey was conducted during the same season as the initial screening with about 85 nonparticipants randomly selected from each organization. Trained interviewers visited the selected member's home as often as necessary to make the contact personally; the number of visits ranged from 1 to 28, with a mean of 5.9. Interviews were scheduled at the member's convenience and conducted primarily in the home; a small number of interviews were conducted at the member's

At the time of the study, Barbara L. Wells was with the Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Md. She is now with the Bureau of Primary Health Care, Health Resources and Services Administration, Rockville, Md. Charles C. Brown is with the Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Md. John W. Horm is with the Division of Health Interview Statistics, National Center for Health Statistics, Hyattsville, Md. Richard A. Carleton is with the Division of Cardiology, Memorial Hospital of Rhode Island, Pawtucket, and with the Department of Medicine, Brown University, Providence, RI. Thomas M. Lasater is with the Division of Health Education, Memorial Hospital of Rhode Island, and with the Department of Community Health, Brown University.

Requests for reprints should be sent to Barbara L. Wells, PhD, Office of Evaluation, Analysis and Research, Bureau of Primary Health Care, Health Resources and Services Administration, Room 7-3A2, E/W Towers, 5600 Fishers Lane, Rockville, MD 20857.

This paper was accepted May 19, 1993.

Editor's Note. See related editorial by Zapka (p 12) in this issue.

TABLE 1—Sociodemographic and Health Characteristics of Religious-Organization Site Screening Participants and Nonparticipants

	Participants (n = 658)	Nonparticipants (n = 326)	Nonparticipants without Previous High Blood Pressure[a] (n = 239)	Nonparticipants with Previous High Blood Pressure[b] (n = 87)
Mean age, y (SD)	45.0 (11.98)	40.6 (12.64)	39.9 (12.75)	42.4 (12.23)
Body mass index, %				
Females >27.3	26.9	29.0	24.2	41.2
Males >27.8	36.2	36.4	32.7	47.2
Diastolic blood pressure, %				
≥90 mmHg	21.9	9.2	12.5	0.0
<90 mmHg	78.1	90.8	87.5	100.0
Worship attendance, %				
Weekly	50.8	32.3	29.7	39.5
1 to 3 times per month	29.8	25.9	25.5	26.8
Occasionally/never	19.4	41.8	44.8	33.7
Distance of residence from religious organization <5 miles, %	94.5	88.6	87.6	91.8
Currently smoke, %	28.1	43.3	46.0	35.6
Cardiovascular disease risk factor knowledge, mean (SD)	3.6 (1.24)	3.3 (1.22)	3.4 (1.24)	3.3 (1.18)
Previously told of high blood pressure, %	31.8	35.6	12.1	100.0

[a]Excludes those with previous high blood pressure and no elevated blood pressure measurement during screening.
[b]Includes those with previous high blood pressure and no elevated blood pressure measurement during screening—that is, a second blood pressure measured systolic <140 mm Hg *and* diastolic <90 mm Hg during screening.

TABLE 2—Predictors of Religious-Organization Site Screening Participation: Multiple Logistic Regression Analysis

	Odds Ratio	95% Confidence Interval
Religious organization activity: worship attendance frequency (0 = rarely, 1 = 1 to 3 times per month, 2 = weekly)	1.63[a]	1.4, 2.0
Convenience: distance of residence from religious organization (0 = <5 miles, 1 = >5 miles)	0.48[**]	0.28, 0.82
Health characteristics		
Currently smoke (0 = no, 1 = yes)	0.61[**]	0.45, 0.83
Body mass index (height/weight^2)	0.51[a,*]	0.36, 0.73
Previously told of high blood pressure (0 = no, 1 = yes)	0.48[a]	0.33, 0.68
Cardiovascular disease risk factor knowledge (0 to 6)[d]	1.20[**]	1.05, 1.3
Diastolic blood pressure (mmHg)	1.90[b,*]	1.6, 2.3
Age	1.20[c,**]	1.05, 1.4

[a]Odds ratio associated with an increase of 5 body mass index units.
[b]Odds ratio associated with an increase of 10 mm Hg in diastolic blood pressure.
[c]Odds ratio associated with an increase of 10 years of age.
[d]0 = 0 risk factors identified, 1 = 1 risk factor identified, 2 = 2 risk factors identified etc.
*$P < .0001$; **$P < .01$.

place of business. The interview and physiological measurements were the same as those used at the religious organization sites. All 658 respondents to the initial screenings in these 10 religious organizations were compared with the nonparticipants.

Logistic regression[16] was used to relate initial screening nonparticipation to 39 factors determined from the interview and physiological measures. These factors fell into seven categories, including sociodemographic characteristics (gender, education, employment status, income); cardiovascular disease physiological risk factors (body mass index, cholesterol levels, systolic and diastolic blood pressure measures); respondent's activity in the religious organization (length of association, committee membership, spouse association); cardiovascular disease risk factor–related behaviors (exercise frequency, current smoking status); and cardiovascular disease knowledge and beliefs (accurate systolic and diastolic blood pressure knowledge, self-rating of cardiovascular disease risk).

A series of increasingly stringent regression analyses[17] were conducted in four stages. First, separate analyses were conducted for each of six predictor variable categories; sociodemographic variables were not analyzed at this stage. Statistically significant categories ($P < .05$) were then included in the second stage to determine the best combination of predictors. In the third stage, the significant predictors were combined with sociodemographic variables. The final stage determined the combination of significant predictors ($P < .05$) that accounted for the largest amount of variation in screening participation.

Results

Of 3277 eligible subjects randomly selected for the initial screening from the membership lists of 20 religious organizations, 1167 agreed to participate. Participation rates in the individual organizations ranged from 17% to 43%. We randomly selected 847 persons who refused the initial screening to request participation in the second screening. Those who moved, died, were no longer members of the religious organization, or could not be found totaled 171 (20%). Of 676 persons contacted, 326 (48%) agreed to participate; the most commonly reported reasons for refusing the second screening were a too-busy schedule or a schedule conflict (18%), or disinterest (15%).

Sociodemographic and health characteristics of initial screening participants and nonparticipants are presented in Table 1. Because these factors and others are interrelated, multivariate logistic regression was used to determine the statistically important factors. The final regression model composition is given in Table 2. Significant determinants of participa-

tion in the initial religious organization screening fall into three primary categories: (1) religious organization activity as measured by worship service attendance; (2) convenience as measured by distance from the screening site; and (3) health characteristics as measured by current smoking status, body mass index, past high blood pressure, cardiovascular disease risk factor knowledge, current diastolic blood pressure, and age. No significant interactions were found between these factors and either age or gender, nor between current blood pressure level and past high blood pressure status. No variables from the categories of social context (e.g., supportiveness of family and friends) or preventive health behaviors and beliefs (e.g., frequency of physical examinations, self-rated health) were found to be statistically significant in the final model.

Discussion

Cardiovascular disease risk factor screening in religious organizations appeared to appeal more to older members who were more active in and resided closer to the organization. Religious organization screenings also attracted more participants with elevated blood pressure readings, 55% of whom reported previous high blood pressure. The proportion of participants with elevated measures exceeds the 11% to 12% detected in random sample households of the same age distribution in communities adjacent to the study religious organizations (T. M. Lasater, PhD, personal communication, March 1993).

About 87 (27%) nonparticipants also reported previous high blood pressure, but current diastolic and systolic blood pressure measured within normal range. It is plausible that these individuals had less incentive to participate in a screening program if they were already being successfully treated for hypertension. The final regression model results did not statistically differ when this 27% was excluded from the analysis.

Because cardiovascular disease screening response rates are so varied in secular organizations—for example, they may be as low as 9%[18] or as high as 81%[19] in work sites—we cannot conclude that the rates in religious organizations are better or worse. What is known is that religious organizations may lack precise denominators because membership lists can be incomplete and outdated; ours were rendered reasonably accurate only after an extensive investment in time, particularly among those organizations with large memberships or loose membership criteria.

Screening participation appeared to increase with higher levels of cardiovascular disease risk factor knowledge, supporting the notion that educational/promotional campaigns be a prerequisite to the conduct of religious organization screenings. The prevalent cardiovascular disease risks of reluctant participants (i.e., smoking and higher body mass index) suggest that such persons are likely to require different intervention strategies to encourage screening participation than their more health-conscious counterparts. Thus, if these religious organizations wish to reach reluctant, at-risk persons in their on-site screening catchment, they need to determine the most effective strategies to attract such persons. □

Acknowledgments

This research was supported in part by the National Heart, Lung, and Blood Institute, grant no. HL28919.

The authors wish to extend their gratitude to B. Jean Bishop for her assistance with data collection and editorial reviews of this paper, and to Richard Ristow for his technical contributions to this paper.

An appendix listing definitions and remarks on variables used in this study are available from the authors.

References

1. National Center for Health Statistics. Advance report of final mortality statistics, 1990. *Month Vital Stat Rep.* January 7, 1993;41(7)(suppl). DHHS publication PHS 93–1120.
2. Report of the National Cholesterol Education Program Expert Panel on detection, evaluation and treatment of high blood cholesterol in adults. *Arch Intern Med.* 1988; 148:36–69.
3. The 1988 report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure. *Arch Intern Med.* 1988;148:1023–1038.
4. *Churches as an Avenue to High Blood Pressure Control.* Bethesda, Md: National Heart, Lung, and Blood Institute; 1992. NIH publication 92–2725.
5. *Statistical Abstract of the United States: 1991.* 111th ed. Washington, DC: US Bureau of the Census; 1991.
6. Smith ED. The role of Black churches in supporting compliance with antihypertensive regimens. *Public Health Nurs.* 1989; 6:212–217.
7. Hatch J, Jackson C. North Carolina Baptist church program. *Urban Health.* 1981;10: 70–71.
8. Kong BW, Miller JM, Smoot RT. Churches as high blood pressure control centers. *Am J Public Health.* 1981;71:1173.
9. Kong BW, Miller JM, Smoot RT. Churches as high blood pressure control centers. *J Natl Med Assoc.* 1982;74:920–923.
10. Levin JS. The role of the Black church in community medicine. *J Natl Med Assoc.* 1984;76:477–483.
11. Perry EJ, Williams BJ. The Memphis church-based high blood pressure program. *Urban Health.* 1981;10:69–70.
12. Saunders E, Kong BW. A role for churches in hypertension management. *Urban Health.* 1983;12:49–51, 55.
13. Sutherland MS, Harris GJ, Barber M. Church-based health promotion activities. Presentation at the 120th Annual Meeting of the American Public Health Association; November 1992; Washington, DC.
14. Strogatz DS, Sherman JA, Elliott D, Ramsey D, Cutchin LM, Ibrahim MA. Community coverage in a rural, church-based, hypertension screening program in Edgecombe County, North Carolina. *Am J Public Health.* 1985;75:401–402.
15. Lasater TM, Wells BL, Carleton RA, Elder JP. The role of churches in disease prevention research studies. *Public Health Rep.* 1986;101:125–131.
16. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons; 1989.
17. *SAS/STAT User's Guide, Version 6.* 4th ed. Cary, NC: SAS Institute Inc; 1990;2.
18. Gemson DH, Sloan RP, Messeri P, Goldberg IJ. A public health model for cardiovascular risk reduction: impact of cholesterol screening with brief nonphysician counseling. *Arch Intern Med.* 1990;150: 985–989.
19. Deacon SP. Screening for coronary risk factors in occupational health practice. *Occup Med.* 1991;41:126–128.

Is Cardiopulmonary Resuscitation Training Deleterious for Family Members of Cardiac Patients?

Kathleen Dracup, DNSc, RN, Debra K. Moser, DNSc, RN, Peter M. Guzy, PhD, MD, Shelley E. Taylor, PhD, and Celine Marsden, MN, RN

ABSTRACT

The purpose of the study was to determine the attitudes toward cardiopulmonary resuscitation (CPR) training and subsequent CPR use of 172 CPR-trained family members of cardiac patients. The majority (88.9%) reported positive attitudes. Only 14 (8.1%) reported feeling too responsible for their family member. One hundred and forty-one (81.9%) said that they would perform CPR if required to do so. Family members do not feel unduly burdened by learning CPR, and CPR training should be recommended to families of patients at risk for sudden cardiac death. (*Am J Public Health.* 1994;84:116–118)

Introduction

Patients with heart disease are at high risk for sudden death. Most of these events occur out of the hospital, in the individual's home.[1–3] Thus, family members are an important target for public education efforts related to reducing the time of response to a sudden-death event. Unfortunately, few family members receive cardiopulmonary resuscitation (CPR) instruction.[4] In fact, the majority of physicians and health educators are reticent to recommend CPR training for families of cardiac patients, citing fears that CPR training will be experienced as a heavy burden and that CPR may be performed inappropriately or ineffectively.[5–9] The purpose of our study was to determine CPR training attitudes, practice patterns, and incidence of CPR use in CPR-trained family members of cardiac patients.

Methods

Family members (n = 337) of patients at risk for sudden cardiac death were enrolled as part of a multicenter clinical trial designed to determine the emotional consequences of different types of CPR training. Subjects were selected from one of four hospital sites by the following criteria: (1) age 25–80 years, (2) literate in English, (3) no CPR training within the past 2 years, and (4) living with a relative at risk for sudden death by virtue of cardiac disease. They were randomly assigned to either a control group (n = 99) that did not receive CPR training or to one of three types of CPR training (n = 238). These CPR training groups were CPR training only, CPR training with the addition of an educational segment about reduction of cardiac risk factors, and CPR training followed by a social support intervention. Because there were no differences among the three CPR-trained groups on attitudes, practice, or retraining, the three groups were combined for purposes of this report.

CPR training consisted of one-person CPR taught by a clinical nurse specialist who was certified as a Basic Cardiac Life Support instructor. CPR was taught according to American Heart Association standards. Classes were structured to optimize learning and retention of CPR skills and knowledge.[10] To further promote retention, a CPR self-study retention packet for home use was mailed to subjects 3 and 6 months after CPR training.[11] Each packet consisted of the following: (1) a booklet outlining the steps in CPR and the components of properly performed CPR; (2) a poster illustrating the steps of CPR; (3) a paper pillow case with a torso imprinted on it that, when placed on a pillow, allows practice of hand placement for pulse check and cardiac compressions; and (4) guidelines for use of the packet.

Of the 238 subjects who had learned CPR, 172 (72%) were contacted a mean of 21 ± 6 months after receiving CPR training. Twenty-eight percent of the subjects were lost to follow-up because they had moved, had changed telephone numbers, were unavailable during the follow-up period because of vacation or business, or refused to answer the survey. Subjects were 59 ± 11 years old and had a mean of 14 ± 3 years of formal education. The majority were female (85%) and spouses (98%).

A registered nurse contacted subjects by telephone and administered a structured interview to collect the following information: (1) CPR use by family members, (2) patient mortality, (3) attitudes of family members toward CPR training, and (4) CPR practice and retraining patterns and confidence in one's ability to perform adequate CPR. To identify changes in CPR knowledge over time and to explore the potential relationship between knowledge of CPR and confidence in one's abil-

Kathleen Dracup, Debra K. Moser, and Celine Marsden are with the School of Nursing, Celine Marsden and Peter M. Guzy are with the School of Medicine, and Shelley E. Taylor is with the Department of Psychology, University of California, Los Angeles.

Requests for reprints should be sent to Kathleen Dracup, DNSc, RN, UCLA School of Nursing, 10833 LeConte Ave, Los Angeles, CA 90024.

This paper was accepted April 9, 1993.

ity to perform effective CPR, we assessed CPR knowledge 2 weeks and 6 months after training. A 24-item, self-administered questionnaire was used (possible scores ranged from 0 to 24). The multiple-choice items were drawn from the national American Heart Association written CPR test.

Results

Four CPR-trained family members used CPR, three successfully and one unsuccessfully. All instances of use were judged by paramedics to be appropriate. Eleven of the 172 (6.4%) cardiac patients whose family member learned CPR died during the follow-up period. Six patients died of cardiac causes, whereas the remaining five died of cancer and other noncardiac causes. In the patients who died of cardiac-related causes, only one death occurred out of the hospital, and CPR was not performed because the CPR-trained family member was not with the victim at the time of his collapse.

The majority of family members reported positive attitudes toward receiving CPR training (see Table 1). The subjects who reported ambivalent feelings toward CPR training worried that they would have to use CPR and would be unable to perform it successfully. Only five subjects reported negative attitudes toward CPR training, largely based on feeling burdened by a sense of responsibility for the survival of their spouse in the case of a cardiac arrest.

The majority said that they would perform CPR if necessary. A small minority would not or were unsure, citing a lack of confidence. The subjects reporting that they would only perform CPR on their family member cited fear of the acquired immunodeficiency syndrome and lawsuits as reasons.

Of the 172 CPR-trained subjects, only 57 (33.1%) subjects reported practicing with the home-use practice packet. Only 14 (8%) subjects received retraining after their initial CPR training. Nonetheless, at the time of the survey, 134 (77.9%) felt that they could still perform adequate CPR.

There was a significant difference in the scores on the questionnaire testing CPR knowledge that was completed at 6 months after CPR instruction; subjects who practiced with the home use practice packet had a mean score of 19.1 ± 2.6, whereas those who did not practice had a mean score of 15.4 ± 2.7 ($P = 0.005$). The 6-month follow-up scores of subjects who practiced were similar to the 2-week follow-up scores of CPR-trained subjects (79% correct vs 80% correct). The subjects who did not practice achieved scores at 6-month follow-up that were similar to those of control subjects who had never been trained in CPR (64% correct vs 62% correct). To assess the relationship between CPR knowledge and subsequent confidence in one's ability to perform CPR, subjects were divided into two groups (i.e., high and low scorers) based on a median split of the CPR test score. Subjects who scored higher on the CPR knowledge test expressed significantly more confidence in their ability to perform adequate CPR ($\chi^2 = 7.08$, degrees of freedom = 2, $P = 0.03$).

Eighty-seven of the interviewed subjects had not learned CPR as part of the initial study. When queried on telephone follow-up, 16 (18.4%) initially untrained family members subsequently learned CPR in a community CPR training class. The 71 family members who did not learn CPR were asked their reasons for not pursuing CPR training. Five did not learn CPR because their spouse died shortly after the study ended. The remaining 66 untrained subjects described a number of barriers to learning CPR, and these are presented in Table 2. The most common reasons for not pursuing CPR training were lack of time, poor health status (of the family member, not the patient), and inattention.

TABLE 1—Attitudes of Family Members of Cardiac Patients 21 Months After Cardiopulmonary Resuscitation (CPR) Training (n = 172)

	No.	%
Feelings about learning CPR		
Very positive	91	52.9
Positive	62	36.0
Ambivalent	7	4.1
Negative	3	1.7
Very negative	2	1.2
No response	7	4.1
Sense of responsibility conferred by learning CPR		
Too much	14	8.1
Right amount	124	72.1
None or very little	34	19.8
Would subject use CPR if needed?		
Yes	141	81.9
No	8	4.7
Maybe	14	8.1
Only on family member	7	4.1
No response	2	1.2

TABLE 2—Reasons Given by Family Members Who Chose Not to Learn CPR (n = 66)

Reasons Cited	No.	%
Lack of time	27	42.9
Did not occur to subject to learn	13	20.6
Medical problems	9	14.3
Inconvenient or too costly	5	7.9
Lack of information about class	3	4.8
Fearful about responsibility	2	3.2
Doctor or nurse did not recommend it	2	3.2
Fear of the acquired immunodeficiency syndrome	1	1.6
Confident in emergency medical system	1	1.6

Note. Not all family members offered a reason, and several offered more than one. Percentages are based on the number of responses given (n = 63).

Discussion

It is important that the family members of those individuals at high risk for sudden death be trained in CPR. Although the average sudden-death incidence is approximately 5% annually in most cardiac populations (e.g., after uncomplicated myocardial infarction),[12] it can reach as high as 50% in patients with advanced congestive heart failure.[13] Unfortunately, families of cardiac patients appear to be the least likely to be trained in CPR.[4,14] Our results also suggest that family members do not seek training on their own. Subjects provided a variety of reasons for not learning CPR, and almost none of these explanations involved fear of too much responsibility or fear of not being able to learn CPR. Although poor health status of the family member proved to be a significant impediment to learning CPR for some, the majority of family members simply chose not to attend to the need for training. This inattention can be addressed by a physician's referral and community education programs, particularly those that target senior citizens and family members of cardiac patients for CPR training.

Many physicians believe in the importance and appropriateness of CPR training for family members of patients with coronary disease, yet fewer than 50% (and in one study[7] fewer than 10%) of physicians recommend CPR training to the family members of their cardiac patients.[5,6] In hypothesizing why physicians' behaviors are at odds with their be-

liefs, Goldberg[9] postulated that physicians believe their patients' families will not understand the importance of CPR training or fear that family members will be unduly frightened if the physician discusses the possibility of sudden death and recommends CPR training. The reluctance of health professionals to recommend CPR training based on such fears, however, may be inappropriate. Once trained in resuscitation techniques, families in the current study did understand the importance of learning CPR and held positive attitudes about having learned it. CPR-trained family members of cardiac patients did not feel that CPR training conferred excessive responsibility. The majority of subjects expressed willingness to use CPR when needed. Those CPR-trained family members who had been called on to use their CPR skills were able to use them successfully and appropriately.

The majority of subjects in the current study did not practice CPR after receiving training and very few retrained in a community CPR course, yet most felt they could perform adequate CPR if called on to do so. Subjects who practiced achieved a score of 80% on a CPR knowledge test at 6 months, whereas subjects who did not practice had a mean score of 64%. Although it is not known how these differences in knowledge might translate to differences in CPR skills in an emergency, the scores of the group who did not practice were similar to those of family members who never learned CPR. These findings indicate that CPR knowledge deteriorates in families, just as it does in health care professionals, and that families do not realize the importance of practice or retraining for maintenance of learned skills and do not appreciate the deterioration of CPR skills that occurs without appropriate reinforcement.[11,15] Clearly, families need to understand the importance of practice and retraining.

In summary, health care professionals should recommend CPR training to the majority of family members of patients at risk for sudden cardiac death. Once trained, families need to be encouraged to retrain annually. The majority of family members do not feel overburdened by training and fears in this regard are unjustified. □

Acknowledgments

This study was funded by a grant from the National Heart, Lung and Blood Institute, National Institutes of Health (RO1 HL32171).

The authors thank the following nurses who interviewed family members: Sharon Coleman, Maria Pappas, and Ruth Ann Kuiper.

References

1. Amey BD, Harrison EE, Straub EJ. Sudden cardiac death: a retrospective and prospective study. *J Am Coll Emerg Physicians.* 1976;5:429–433.
2. Bergner L, Eisenberg MS. CPR: who should be trained? *Emerg Health Serv Q.* 1982;1:41–48.
3. Goldstein S, Friedman L, Hutchinson R, et al. Timing, mechanism, and clinical setting of witnessed death in post myocardial infarction patients. *J Am Coll Cardiol.* 1984; 3:1111–1117.
4. Pane GA, Salness KY. A survey of participants in a mass CPR training course. *Ann Emerg Med.* 1987;16:1112–1116.
5. Mandel LP, Cobb LA, Weaver WD. CPR training for patients' families: do physicians recommend it? *Am J Public Health.* 1987;77:727–728.
6. St. Louis P, Carter WB, Eisenberg MS. Prescribing CPR: a survey of physicians. *Am J Public Health.* 1982;72:1158–1160.
7. Goldberg RJ, DeCosimo D, St. Louis P, Gore JM, Ockene JK, Dalen JE. Physicians' attitudes and practices toward CPR training in family members of patients with coronary heart disease. *Am J Public Health.* 1985;75:281–283.
8. Nelson KM. Cardiopulmonary resuscitation training for families of cardiac patients. *Cardiovasc Nurs.* 1979;6:28–32.
9. Goldberg RJ. Physicians and CPR training in high-risk family members. *Am J Public Health.* 1987;77:671–672. Editorial.
10. Dracup K, Heaney D, Taylor SE, Guzy PM, Breu C: Can family members of high-risk cardiac patients learn CPR? *Arch Intern Med.* 1989;149:61–64.
11. Moser DK, Dracup K, Guzy PM, Taylor SE, Breu C. Cardiopulmonary resuscitation skills retention in family members of cardiac patients. *Am J Emerg Med.* 1990; 8:498–503.
12. Bardy GH. Sudden cardiac death. In: Zipes DP, Rowlands DJ, eds. *Progress in Cardiology.* Philadelphia, Pa: Lea & Febiger; 1988.
13. Franciosa JA, Wilen M, Ziesche S, Cohn JN. Survival in men with severe chronic left ventricular failure due to either coronary artery disease or idiopathic dilated cardiomyopathy. *Am J Cardiol.* 1983;51: 831–836.
14. Goldberg RJ, Gore JM, Love DG, Ockene JK, Dalen J. Layperson CPR—are we training the right people? *Ann Emerg Med.* 1984;13(part 1):701–704.
15. Kaye W, Mancini ME. Retention of cardiopulmonary resuscitation skills by physicians, registered nurses, and the general public. *Crit Care Med.* 1985;14:620–622.

Educating Patients with Limited Literacy Skills: The Effectiveness of Printed and Videotaped Materials about Colon Cancer

Cathy D. Meade, PhD, RN, W. Paul McKinney, MD, and Gary P. Barnas, MD

ABSTRACT

We investigated whether printed or videotaped information is more effective in enhancing colon cancer knowledge. Subjects (n = 1100) were randomized into three groups: to receive a booklet, view a videotape, or receive no intervention. Subjects receiving the intervention showed increased knowledge compared with control subjects (booklet = 23% and videotape = 26% vs no intervention = 3%). Findings suggest that personalized educational materials are effective in enhancing colon cancer knowledge. (*Am J Public Health.* 1994;84:119–121)

Introduction

Colon cancer, the second most common malignancy among men and women in the United States, results in over 61 000 deaths each year.[1] Efforts to reduce colon cancer mortality and morbidity focus primarily on early detection and treatment, given the high survival rate in early stages of the disease.[2–5] Printed materials are commonly used to communicate screening guidelines and detection practices, yet they are often produced at reading levels above that of the intended reader.[6–9] For those with low reading skills, videotapes may offer a significant advantage over printed materials because of their visual appeal. Videotaped instruction has been demonstrated to be as effective as other instructional methods and often more effective than printed materials alone.[10–12] Though the efficacy of written and videotaped methods has been widely reported,[10,13–15] we believe this is the first study to compare the effect of printed vs videotaped colon cancer information on enhancing knowledge among individuals with limited reading skills.

Methods and Measures

Pilot data from 85 clinic patients revealed that most did not recognize that colon cancer was common, were not familiar with American Cancer Society screening guidelines, and felt that a colostomy was necessary when colon cancer was detected. A disparity between stated median educational level (grade 10) and the median reading level as estimated by the Wide-Range Achievement Test II[16] (grade 6) confirmed earlier results.[17]

Instruments

The instruments were produced by the investigators and were based on current cancer literature, a panel of experts, and our pilot data.

Booklet. Bold advance headers introduced the reader to the five sections of the 8½ × 17-inch trifold booklet: facts about colon cancer, facts about the colon and rectum, signs and symptoms of colon cancer, early detection of colon cancer, and a summary. *Colon Cancer: Early Detection Can Save Your Life* was printed in 12-point type with 14-point headers and written at a grade 5–6 reading level.[18,19]

Videotape. Content of the videotape mirrored that of the booklet. Similar headings were placed in the videotape to create visual cues for the viewer. The videotape was filmed in our clinic and showed familiar surroundings. The incorporation of modeling[11] was evidenced by showing patients participating in desired screening behaviors. The videotape was 7½ minutes in length, which approximated the time required to read the booklet.

Pretest/posttest. To evaluate colon cancer knowledge and recall, 24 questions written at grade 5–6 reading level were developed by the investigators and evaluated by a panel of experts. Cronbach's internal consistency alpha coefficient was calculated (r = .63). Questions were systematically divided into two, 12-question tests based on question content, reliability measures, item difficulty, and logical pairing of questions. Each test was comprised of eight true or false and four multiple-choice questions.

Study Population and Data Collection

Subjects (n = 1100) were selected from the Primary Care Clinic at the Milwaukee County Medical Complex, a ma-

Cathy D. Meade is with the School of Nursing, University of Wisconsin–Milwaukee, Milwaukee. W. Paul McKinney is with the Department of Internal Medicine, University of Texas Southwestern Medical School, Dallas. Gary P. Barnas is with the Department of Internal Medicine, Medical College of Wisconsin, Milwaukee.

Requests for reprints should be sent to Cathy D. Meade, PhD, RN, University of Wisconsin–Milwaukee, School of Nursing, PO Box 413, Milwaukee, WI 53201.

This paper was accepted June 23, 1993.

Editor's Note. See related editorial by Zapka (p 12) in this issue.

TABLE 1—Summary of Scores, by Group

	Intervention Groups		
	Control (n = 356)	Experimental-Booklet (n = 370)	Experimental-Videotape (n = 374)
Pretest (range: 1, 12)			
Mean	7.5	7.5	7.4
95% confidence interval	7.3, 7.7	7.3, 7.7	7.2, 7.6
Posttest (range: 0, 12)			
Mean	7.7	9.2	9.3
95% confidence interval	7.5, 7.9	9.0, 9.4	9.1, 9.5
Score difference (range: −5, 10)			
Mean	0.2	1.7	1.9
95% confidence interval	0.0, 0.4	1.4, 1.9	1.7, 2.2
Score improvement (post-pre)/(pre)	3%	23%	26%

TABLE 2—Mean Scores, by Medium (Booklet or Video) and Reading Skill Level

Medium	Reading Skill[a]	No.	Scores (mean ± SD)		Mean Score Difference	95% Confidence Interval
			Pretest	Posttest		
Booklet	Low	184	7.05 ± 1.72	8.6 ± 1.73	1.54	1.23, 1.85
Booklet	High	185	8.05 ± 1.83	9.85 ± 1.60	1.79	1.49, 2.09
Videotape	Low	216	6.89 ± 1.99	8.68 ± 1.79	1.79	1.45, 2.11
Videotape	High	158	8.01 ± 1.60	10.15 ± 1.68	2.13	1.82, 2.44

[a]Low = Wide-Range Achievement Test II score below grade 7. High = Wide-Range Achievement Test II score at grade 7 or higher.

jor teaching affiliate of the Medical College of Wisconsin. Selection criteria included age of 50 years or older, ability to speak and read English, absence of visual and hearing impairments, ability to give free consent, and eligibility for at least one colon cancer screening measure within the recommended interval. Subjects were allocated randomly by the permuted block method into one of three groups: control, booklet, or videotape. All subjects were first given the pretest. Control group subjects received no educational intervention. Other subjects either were asked to read the booklet or view the videotape. All subjects were given the posttest and the Wide-Range Achievement Test II[17] and were questioned about their demographic background. SPSS-X computer programs were used to perform the analyses.

Results

There were no significant differences in age (mean = 60.6 years), race (54% Black, 44% White), sex (72% female), and Wide-Range Achievement Test II Scores among the groups. The median level of reported education was grade 11, and the median reading level estimated by the Wide-Range Achievement Test II was grade 7, a difference of 4 grades. With a one-way analysis of variance, no significant differences were found in the pretest scores among the groups (Table 1). Tukey's Honestly Significant Difference analysis revealed that the difference in posttest minus pretest score between each intervention group and the control group was large enough to be significant at the .05 level of confidence. No statistically significant difference was noted between the booklet and videotape groups.

Subjects in both experimental groups were then divided into four groups according to their Wide-Range Achievement Test II scores (Table 2). Those with grade 7 or higher scores were considered to have high reading skills, whereas those with lower than grade 7 scores were considered to have low reading skills. This definition was based on their expected reading ability within 2 years of the reading level of the booklet. No statistically significant differences in score improvements were found among the four subgroups according to analysis of variance.

Discussion

Both printed and videotaped materials enhanced colon cancer knowledge among patients with limited literacy skills. One possible explanation for this finding is that the interventions were tailored to our target group, with special attention given to developing content relevant to their learning needs, designing the instruments to reflect ethnic diversity, organizing content in a clear manner, using the active voice, writing or narrating in a conversational style, using short words and sentences, incorporating headers and cues, summarizing points, and pretesting the tools.[20–26] The effectiveness of both educational media in enhancing patient knowledge also may have been influenced by consideration of our patients' attitudes towards colon cancer and associated procedures,[27] as well as by the use of the pretest. Our findings support the hypothesis that printed materials written at low reading levels (grade 5–6) can be effective substitutes in clinics without access to expensive audiovisual equipment.[28–30] We recommend that formative and summative evaluations of educational materials be done to ensure that they are acceptable, appropriate, and comprehensible to the target group.[6,20,24]

The inference from our data that both interventions enhanced patients' knowledge about colon cancer should be viewed with caution because only short-term knowledge recall was evaluated. Future studies will be required to determine how long this knowledge is retained and its effect on attitudinal and behavioral outcomes. □

Acknowledgments

This work was supported by a grant from the Wisconsin Department of Health and Social Services, Bureau of Community Health and Prevention.

An earlier version of this paper was presented at the annual meeting of the American Public Health Association, November 10–11, 1991, Atlanta, Ga.

We express our appreciation to Thuy Lindsey and Shawneen Schmitt, RN, for their ongoing assistance and contributions to this project.

References

1. American Cancer Society. *1992 Cancer Facts and Figures*. New York, NY: American Cancer Society: 1992.
2. *Surveillance, Epidemiology, and End Results. Incidence and Mortality Data: 1973–1977*. Bethesda, Md: National Cancer Institute; 1981. NCI Monograph 57.
3. Gullatte MM. Cancer prevention and early detection in Black Americans: colon and

rectum. *J Natl Black Nurse Assoc.* 1989; 3:49–56.
4. Gupta TP, Silverman AL, Desai TK, Luk GD. Prevention of colon cancer in primary care practice. *Prev Pract.* 1989;16:157–171.
5. Adams M. Information and education across the phases of cancer care. *Semin Oncol Nurs.* 1991;7:105–111.
6. Redman BK. *The Process of Patient Education.* 6th ed. St. Louis, Mo: CV Mosby Co; 1988.
7. Boyd MD, Feldman RHL. Health information seeking and reading and comprehension abilities of cardiac patients. *J Card Rehabil.* 1984;4:343–347.
8. Swanson JM, Forest K, Ledbetter C, Hall S, Holstine EJ, Shafe MR. Readability of commercial and generic contraceptive instructions. *Image J Nurs Scholar.* 1990;22: 96–100.
9. Meade CD, Diekmann J, Thornhill D. Readability of American Cancer Society patient education literature. *Oncol Nurs Forum.* 1992;19:51–55.
10. Gagliano ME. A literature review on the efficacy of video in patient education. *J Med Educ.* 1988;63:785–792.
11. Nielsen E, Sheppard MA. Television as a patient education tool: a review of its effectiveness. *Patient Educ Counseling.* 1988;11:3–16.
12. Uzark K, Rosenthal A, Behrendt D, Becket M. Use of videotape to promote parenting of infants with serious congenital heart defects. *Patient Educ Counseling.* 1985;7:111–119.
13. Li VC, Coates TJ, Spielberg LA, Ewart CK, Dorfman S, Huster WJ. Smoking cessation with young women in public family planning clinics: the impact of physician messages and waiting room media. *Prev Med.* 1984;13:477–489.
14. Manderino MA, Bzked VM. Effects of modeling and information on reactions to pain: a childbirth-preparation analogue. *Nurs Res.* 1984;33:9–14.
15. Marshall WR, Rothenberger LA, Bunnell SL. The efficacy of personalized audiovisual patient education material. *J Fam Pract.* 1984;19:659–663.
16. Jastak S, Wilkinson GS. *The Wide-Range Achievement Test: Revised Administration Manual.* Wilmington, Del: Jastak Associates; 1984.
17. Meade CD, Byrd JC. Patient literacy and the readability of smoking education literature. *Am J Public Health.* 1989;79:204–206.
18. LACE Network. *Minnesota Interactive Readability Approximation Program: Users Guide.* LaCrosse, Wis: University of Wisconsin–LaCrosse Computer Center; 1981.
19. McLaughlin HG. SMOG-grading: a new readability formula. *J Reading* 1969;12: 641–646.
20. Doak CC, Doak LG, Root JH. *Teaching Patients with Low Literacy Skills.* New York, NY: JB Lippincott; 1985.
21. Villejo L, Meyers C. Brain function, learning styles, and cancer patient education. *Semin Oncol Nurs.* 1991;7:97–102.
22. Duffy MM. Selecting educational materials for patients with limited reading abilities. *Am Nephrology Nurs Assoc J.* 1988;15: 114–117.
23. Nitzke S. Reaching low-literate adults with printed nutrition materials. *J Am Diet Assoc.* 1987;87(suppl):73–77.
24. *Making Health Communications Work: A Planner's Guide.* Bethesda, Md: National Cancer Institute Office of Cancer Communications; 1992. NIH Publication 92-1493.
25. Weinrich SP, Blesch KS, Dickson GW, Nussbaum JS, Watson EJ. Timely detection of colorectal cancer in the elderly. *Cancer Nurs.* 1989;12:170–176.
26. Hussey LC. Overcoming the clinical barriers of low literacy. *J Gerontol Nurs.* 1989; 17:27–29.
27. American Cancer Society. A summary of the American Cancer Society report to the nation: cancer in the poor. *Cancer J Clin.* 1989;39:263–287.
28. Baker GC, Newton DE, Bergstresser PR. Increased readability improved comprehension of written information for patients with skin disease. *J Am Acad Dermatol.* 1988;19:1135–1141.
29. Meade CD, Byrd JC, Lee M. Improving patient comprehension of literature on smoking. *Am J Public Health.* 1989;79: 1411–1412.
30. Owen PM, Porter KA, Frost CD, O'Hare E, Johnson E. Determination of the readability of educational materials for patients with cardiac disease. *J Cardiopulmonary Rehabil.* 1993;13:20–24.

Call for Nominations for 1994 Public Health Nutrition Awards

The Food and Nutrition Section of the American Public Health Association would like to honor individuals who have made outstanding contributions to nutrition within public health. Nominations are being solicited for the Mary C. Egan Award, the Catherine Cowell Award, and the Excellence in Dietary Guidance Award. The Catherine Cowell Award is presented to an individual who is innovative and who has a keen vision in developing programs in public health nutrition. The Mary C. Egan Award recognizes professional contribution to and special service in public health nutrition.

Nominations must be received by *May 1, 1994.* For more information contact Yvonne Bronner, PhD, RD, Johns Hopkins University, School of Hygiene and Public Health, Department of Maternal and Child Health, 624 N Broadway, Room 245, Baltimore, MD 21205; tel (410) 955-3804.

A2623-1

American Journal of Public Health

February 1994, Vol. 84, No. 2 Established 1911

Editorials

173 Health Care Reform and Public Health: Weighing the Proposals *M. Susser*
175 AIDS—Displaying the Global Dynamics

Annotations

178 The Need to Mobilize Support for the Wellstone-McDermott-Conyers Single-Payer Proposal *V. Navarro*
180 Alcohol Advertising and Public Health: An Urgent Call for Action *J. F. Mosher*

Public Health Policy Forum

182 Should Single-Payer Advocates Support President Clinton's Proposal for Health Care Reform? *E. R. Brown*
186 Commentary: Recent Reforms in the British National Health Service—Lessons for the United States *W. W. Holland and C. Graham*

Articles

VARIOUS TOPICS

190 Factors Influencing the Duration of Work-Related Disability: A Population-Based Study of Washington State Workers' Compensation *A. Cheadle, G. Franklin, C. Wolfhagen, J. Savarino, P. Y. Liu, C. Salley, and M. Weaver*
197 Circumcision and Sexually Transmitted Diseases *L. S. Cook, L. A. Koutsky, and K. K. Holmes*

MENTAL HEALTH/SUBSTANCE ABUSE

202 Employment Status and Depressive Symptoms in Middle-Aged Women: A Longitudinal Investigation *J. T. Bromberger and K. A. Matthews*
207 Providing Medical Care to Methadone Clinic Patients: Referral vs On-Site Care *A. Umbricht-Schneiter, D. H. Ginn, K. M. Pabst, and G. E. Bigelow*
211 The Relationship between Mortality and Intensity of Inpatient Alcoholism Treatment *J. Y. Bunn, B. M. Booth, C. A. L. Cook, F. C. Blow, and J. C. Fortney*
215 Completion Rates of Clients Discharged from Drug and Alcohol Treatment Programs in Washington State *T. Wickizer, C. Maynard, A. Atherly, M. Frederick, T. Koepsell, A. Krupski, and K. Stark*
222 Ethnicity and the Use of Outpatient Mental Health Services in a National Insured Population *D. K. Padgett, C. Patrick, B. J. Burns, and H. J. Schlesinger*
227 Mental Health Status as a Predictor of Morbidity and Mortality: A 15-Year Follow-Up of Members of a Health Maintenance Organization *T. Vogt, C. Pope, J. Mullooly, and J. Hollis*
232 Dementia-Free Life Expectancy in France *K. Ritchie, J.-M. Robine, L. Letenneur, and J.-F. Dartigues*
237 The Prevalence and Demographics of Pathological Gamblers: Implications for Public Health *R. A. Volberg*
242 Sedatives and Hypnotics in Stockholm: Social Factors and Kinds of Use *G. Blennow, A. Romelsjö, H. Leifman, A. Leifman, and G. Karlsson*
247 Societal-Level Predictors of Groups' Drinking Patterns: A Research Synthesis from the Collaborative Alcohol-Related Longitudinal Project *K. M. Fillmore, J. M. Golding, E. V. Leino, C. R. Ager, and H. P. Ferrer, with the Collaborative Alcohol-Related Longitudinal Project*
254 Television Beer Advertising and Drinking Knowledge, Beliefs, and Intentions among Schoolchildren *J. W. Grube and L. Wallack*

(continued page 161)

APHA
American Public Health Association

Featuring Mental Health/ Substance Abuse

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

February 1994, Vol. 84, No. 2 ISSN: 0090-0036 Established 1911

260 Emotional and Behavioral Problems and Severe Academic Delays among Sheltered Homeless Children in Los Angeles County *B. T. Zima, K. B. Wells, and H. E. Freeman*

265 Risk Factors for Homelessness among Schizophrenic Men: A Case–Control Study *C. L. M. Caton, P. E. Shrout, P. F. Eagle, L. A. Opler, A. Felix, and B. Dominguez*

271 AIDS, HIV Infection, and Illicit Drug Use within Inner-City Families and Social Networks *A. Pivnick, A. Jacobson, K. Eric, L. Doll, and E. Drucker*

275 Injecting Risk Behavior among Drug Users in Amsterdam, 1986 to 1992, and Its Relationship to AIDS Prevention Programs *E. J. C. van Ameijden, A. (J) A. R. van den Hoek, and R. A. Coutinho*

Public Health Briefs

VARIOUS TOPICS

282 Chronic Fatigue Syndrome: Have Flawed Assumptions Been Derived from Treatment-Based Studies? *J. A. Richman, J. A. Flaherty, and K. M. Rospenda*

MENTAL HEALTH/SUBSTANCE ABUSE

284 The Provision and Use of Mental Health Services in Nursing Homes: Results from the National Medical Expenditure Survey *M. A. Smyer, D. G. Shea, and A. Streit*

287 Psychotic Inpatients' Social Class and Their First Admission to State or Private Psychiatric Baltimore Hospitals *C. Muntaner, P. Wolyniec, J. McGrath, and A. E. Pulver*

290 Psychiatric and Substance Abuse Disorders among Male Urban Jail Detainees *L. A. Teplin*

294 Enhancing the Self-Report of Alcohol Consumption in the Community: Two Questionnaire Formats *A. C. King*

297 The Frequency and Nature of Alcohol and Tobacco Advertising in Televised Sports, 1990 through 1992 *P. A. Madden and J. W. Grube*

300 The Illegal Sale of Cigarettes to US Minors: Estimates by State *K. M. Cummings, T. Pechacek, and D. Shopland*

303 A Pilot Syringe Exchange Program in Washington, DC *D. Vlahov, C. Ryan, L. Solomon, S. Cohn, M. R. Holt, and M. N. Akhter*

Public Health Then and Now

305 Is Liquor Intoxicating? Scientists, Prohibition, and the Normalization of Drinking *P. J. Pauly*

Notes from the Field

314 Using Race-Specific Community Programs to Increase Organ Donation among Blacks *B. A. Hong, D. F. Kappel, M. Whitlock, T. Parks-Thomas, and B. Freedman*

315 Clinica Tibas: A Public–Private Partnership for Health Care Delivery in Costa Rica *P. E. Pezza and J. F. B. Bolanos*

316 Assessing Community Health Concerns and Implementing a Microgrant Program for Self-Help Initiatives *A. Paine-Andrews, V. T. Francisco, and S. B. Fawcett*

Other Journal Departments

162 APHA Masthead

169 Advertisers' Index

170 What *AJPH* Authors Should Know

173 Journal Masthead

319 Letters to the Editor

322 Association News: In Memoriam

325 Book Corner

328 Job Opportunities

APHA

American Public Health Association

American Journal of Public Health

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 Fifteenth Street, NW, Washington, DC 20005; (202) 789-5600. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*

Helen Rodriguez-Trias, MD, *Immediate Past President*

Caswell A. Evans, Jr, DDS, MPH, *President-Elect*

Leonard M. Schuman, MD, *Vice-President (USA)*

John E. F. Hastings MD, MPH, *Vice-President (Canada)*

Oscar Feo, MD, MSc, *Vice-President (Latin America)*

Jay H. Glasser, PhD, *Treasurer*

Allan G. Rosenfield, MD, *Chair, Executive Board*

Shirley F. Randolph, MSPH, *Speaker, Governing Council*

Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*

Carol Easley Allan, PhD, RN (1997)

Caswell A. Evans, Jr, DDS, MPH, *President-Elect*

Patricia P. Evans, MPH (1994)

Eugene Feingold, PhD, JD, *President*

Jay H. Glasser, PhD, *Treasurer*

Alan R. Hinman, MD, MPH (1995)

Margo D. Kerrigan, MPH (1995)

Joanne Lukomnik, MD (1996), *Vice-Chair*

Steven Uranga McKane, DMD, MPH (1995)

Carlos W. Molina, EdD (1997)

Linda Rae Murray, MD, MPH (1996)

Hila Richardson, DrPH, MPH (1996)

Helen Rodriguez-Trias, MD, *Immediate Past President*

Reed V. Tuckson, MD (1994)

Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*

Frances Kissling, *Chair, Action Board*

George E. Pickett, MD, MPH, *Speaker, Program Development Board*

Shirley F. Randolph, MSPH, *Speaker, Governing Council*

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, Inc, 1015 Fifteenth St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health,* 1015 Fifteenth St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 Fifteenth St, NW, Washington, DC 20005. AJPH allotted $40; TNH allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 Fifteenth St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, AJPH Reprints, APHA, 1015 Fifteenth St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The American Journal of Public Health is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

Factors Influencing the Duration of Work-Related Disability: A Population-Based Study of Washington State Workers' Compensation

ABSTRACT

Objectives. The purpose of this study was to examine factors predictive of duration of work-related disability.

Methods. Multivariate survival analysis techniques were used to conduct a population-based, retrospective cohort study on a random sample of 28 473 workers' compensation claims from Washington State filed for injuries occurring in 1987 to 1989. The principal outcome measure was length of time for which compensation for lost wages was paid, used as a surrogate for duration of temporary total disability.

Results. The findings suggest that, even after adjusting for severity of injury, older age, female gender, and a diagnosis of carpal tunnel syndrome or back/neck sprain significantly predict longer duration of disability. Other predictors that were stable and significant, but involved lower magnitudes of effect included divorced marital status, firm size of fewer than 50 employees, higher county unemployment rates, and construction and agricultural work.

Conclusions. Greater disability prevention efforts targeting these higher risk subgroups could have significant economic and public health effects. The greatest impact may be on claimants who remain disabled at 6 months after an injury that did not require hospitalization. (*Am J Public Health.* 1994;84:190–196)

Allen Cheadle, PhD, Gary Franklin, MD, MPH, Carl Wolfhagen, MPA, James Savarino, PhD, P. Y. Liu, PhD, Charles Salley, MS, and Marcia Weaver, PhD

Introduction

Disability associated with work-related injuries is an increasingly serious problem in workers' compensation. Among work-related injuries, the small fraction of disabilities that are of long duration account for a large fraction of workers' compensation costs.[1] More important, if an injured worker has not regained satisfactory work status 2 years after the injury, his or her chances of doing so thereafter are very low.[2] Given the high costs to society of lost productivity and the high human costs of disability to an injured worker, long-duration work-related disability is a serious public health concern.

Although many studies have investigated factors that predict disability after work-related injuries,[3–17] most have been limited to specific types of injury; have used relatively small, non-population-based samples; or have not adjusted for severity of injury. This paper reports on analyses conducted on a random sample of all workers' compensation claims from Washington State filed for injuries occurring in 1987 to 1989. Because Washington is one of six states where employers are required to obtain workers' compensation insurance exclusively through a state fund (or to self-insure), the sample is broadly representative of work-related injuries occurring in the state for that time period. Survival analysis techniques were used to predict duration of disability as a function of injury nature and severity and a variety of worker, firm, and industry characteristics.

Methods

Sampling Frame and Study Design

All workers' compensation claims filed between January 1, 1987, and December 31, 1989, were identified through the extensive claims and medical bill payment databases of the Washington State Department of Labor and Industries. For the purposes of this study, only employees with claims involving at least 4 days of time lost from work (compensable claims) were considered to be at risk for long-term disability (Figure 1). Exclusions from the baseline population were as follows: (1) claims involving 3 or fewer days lost from work as a result of occupational injury or illness (noncompensable claims) and (2) claims from self-insured employers (approximately 350 employers, representing one third of covered workers), whose reporting requirements are inadequate for research purposes.

A random sample of 10 000 compensable claims was drawn in each of the 3 years. Claims with missing information for one or more independent variables (n = 1527; 5.1%) were deleted from the sample, leaving a final incidence cohort of 28 473 claims. A total of 592 claims (2.0%) lacked the worker's monthly wage; an-

Allen Cheadle, James Savarino, and Marcia Weaver are with the Department of Health Services, University of Washington, Seattle. Gary Franklin, Carl Wolfhagen, and Charles Salley are with the Washington State Department of Labor and Industries, Olympia. Gary Franklin is also with the Department of Environmental Health, University of Washington. P. Y. Liu is with the Fred Hutchinson Cancer Research Center, Seattle.

Requests for reprints should be sent to Allen Cheadle, PhD, Department of Health Services, JD-43, University of Washington, Seattle, WA 98195.

This paper was accepted May 11, 1993.

Note. The opinions and conclusions are the authors' and do not necessarily reflect the views of the Washington State Department of Labor and Industries.

other 935 (3.1%) had missing data on other covariates. Duration of disability and most of the demographic variables were not significantly different between included and excluded claims; when there were significant differences, the magnitudes were small. This 26% sample of the total compensable claim population represents the work-related disability experience of two thirds of the nonfederally covered workers in Washington State (approximately 900 000 full-time equivalent workers).

The study involved a population-based, retrospective cohort design[18] (Figure 1), with duration of disability serving as the principal outcome measure. Time-loss data were extracted in June 1991; the mean duration of follow-up for the cohort was 35.4 months. All known or suspected predictors of disability duration (Table 1) were obtained from (1) computerized claim data collected from the initial accident report (gender, age, marital status, dependents, benefit rate [from baseline wage], type of injury, injury date, county of injury, type of firm ownership [public or private], and Standard Industrial Classification code); (2) medical bill payment data (the injury severity marker was hospitalization within 28 days of injury); (3) state unemployment surveys (county unemployment rate); (4) work hours reported by the employer, measured in the number of full-time equivalent workers employed (firm size); and (5) computerized codes relevant to premium rating (experience vs retrospective rating).

The US Department of Labor Z16.2 codes developed by the American National Standards Institute were used in categorizing the nature of injuries. These codes have wide use in workers' compensation systems. Although carpal tunnel injuries involved a body part code of upper extremity, wrist, or hand and were classified as nervous system condition/disease or bursitis, the Department of Labor and Industries specifically coded for this condition during the time period under study.

The state workers' compensation fund has two ratings that adjust the premium costs paid by employers to reflect their accident costs: experience rating and retrospective rating. Time-loss and medical costs are experience rated for all employers based on a 3-year average of past costs. Retrospective rating is an optional incentive system available to state fund employers. Employers with a rating better than the average of those for their job classifications and better than their own past ratings receive a refund of the difference; those whose ratings decline pay the difference to the state. Firms in the retrospective rating program have disability prevention incentives similar to those of the self-insured firms excluded from our sample of claims.

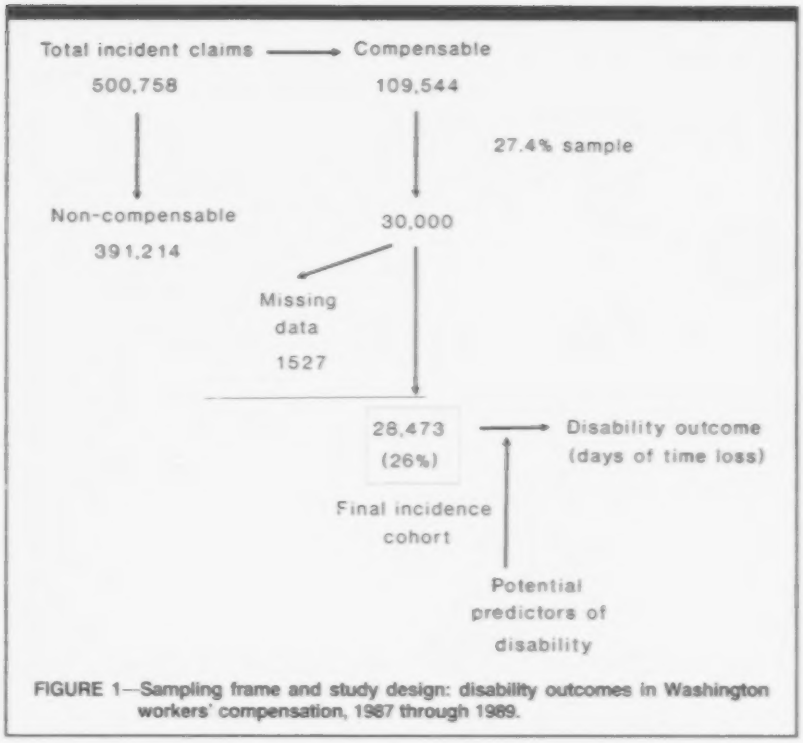


FIGURE 1—Sampling frame and study design: disability outcomes in Washington workers' compensation, 1987 through 1989.

The last set of variables shown in Table 1 was included to capture the effect of the benefit structure on duration of disability. The basic benefit rate in Washington is 60% of the monthly wage at the time of injury. The rate increases on the basis of marital status (workers with a spouse receive an additional 5% benefit) and number of dependents (2% per dependent, up to a maximum of 10%). The maximum benefit rate, regardless of monthly wage, is 75% of the state average weekly wage prior to July 1, 1988, and 100% of the average weekly wage after July 1, 1988. The minimum monthly benefit of roughly $200 also varies on the basis of marital status and number of dependents.

Since the benefit rate for most workers in our sample was determined by their marital status and number of dependents, we could not separate out the independent effect of the benefit rate on return to work. However, for workers with a monthly wage above that yielding the maximum benefit amount (maximum benefit/.6), the benefit rate did not depend on either marital status or number of dependents. In preliminary analyses, we included a dummy variable equal to 1 when the wage was above this maximum amount, as well as a term for the interaction between this dummy variable and the benefit rate. The coefficient for this interaction variable, approximating the "pure" effect of the benefit rate for a select sample of high-income workers, was not statistically significant. Therefore, we used only the simpler specification shown in Table 1 to control for the effect of workers' compensation benefits on duration of disability.

In another preliminary analysis, we linked the 1980 census occupational codes recorded on the claim form to a file containing a variety of job characteristics, including educational requirements, prestige, and income.[19-21] These indirect measures of socioeconomic status were generally insignificant or inconsistent predictors of disability duration; consequently, only the benefit rate variables were included in the final analysis.

All of the variables included in the model are categorical. For continuous variables (age, benefit rate, firm size), the analysis was also performed with a continuous specification (including a quadratic term). In all cases, the results were very similar regardless of specification, and categorical results are reported here for ease of interpretation.

The claim rather than the worker was the unit of analysis, and a small percentage of workers in our sample (2.8%) had mul-

TABLE 1—Definitions and Descriptive Statistics for Independent Variables, 1987–1989 Combined Sample (n = 28 473)

	Employees, %	Mean Duration of Disability, d[a]
Gender		
Male	74	112.7
Female	26	134.9
Age, y		
<30	42	83.5
30–44	41	136.6
45	17	160.8
Marital status		
Never married	35	86.3
Married	50	132.9
Divorced	14	146.6
Widowed	1	169.2
Dependents		
No	66	107.1
Yes	34	141.3
Type of injury[b]		
Fracture	11	126.0
Sprain, back/neck	34	145.6
Sprain, other	20	110.4
Carpal tunnel	2	159.9
All other	34	91.4
Hospitalized within 28 days		
No	94	110.4
Yes	6	257.4
County		
King	35	97.7
Pierce	11	112.0
Snohomish	7	123.8
Spokane	6	157.7
Other: east	22	132.5
Other: west	19	130.6
Year of injury		
1987	34	...[c]
1988	33	...[c]
1989	33	...[c]
Unemployment rate, %		
<5	27	90.3
5–7	36	112.6
>7	37	145.1
Industry[d]		
Agricultural	7	130.2
General construction	2	157.4
Heavy construction	10	125.1
Special trade construction	9	140.0
Wood products	4	131.1
Trucking	7	96.3
Wholesale trade	7	81.1
Restaurant service	11	122.6
Health services	3	105.6
Auto dealers	4	152.9
All other	37	113.2
Firm size		
<50 employees	57	128.0
≥50 employees	43	106.2
Ownership		
Private	95	119.3
Government	5	106.9
Retrospective rating program		
At time of injury		
No	79	122.4
Yes	21	104.6
Participation over 4 years (1986–1989)		
Not participating	70	122.5
In entire period	10	98.1
Came in and stayed	9	102.5
In and out of program	12	124.4
Benefit rate, % of wage		
<60	15	145.7
60–65	46	100.8
>65	39	129.1
Overall	100	118.6

Note. All variables are categorical.
[a]Mean duration of disability (days) for each category, including censored observations, unadjusted for covariates.
[b]Nature of injury and body part involved (from Z16.2 codes).
[c]Not shown because of the effect of censoring on the results.
[d]Standard Industrial Classification codes.

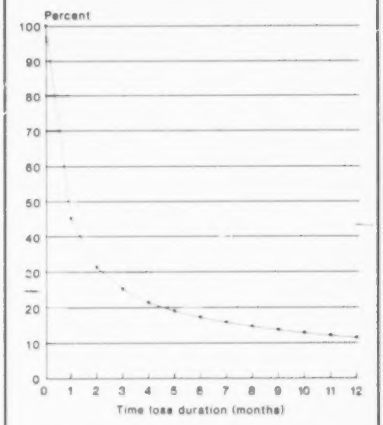


FIGURE 2—Overall percentage of workers receiving disability payments after the indicated number of months since injury.

tiple claims during the 3-year study period. Theoretically, all of these multiple claims, if given a separate identifier, should have involved new injuries and therefore could be regarded as independent events. However, individual characteristics probably influenced the duration of disability for all injuries for a particular worker, however dissimilar the injuries. To examine the impact of these claims on the results, one claim was selected at random from all workers with more than one claim, and the analyses reported below were rerun. The results were virtually identical to those reported here, and it was concluded that any potential nonindependence effect due to multiple claims was relatively minor.

Statistical Analysis

Survival analysis techniques were used to model duration of disability as a function of the independent variables. Cox regression models,[22,23] which make relatively few assumptions about the shape of survival curves, were used in these analyses. The principal assumption of the Cox model is that the relative "hazard" (in this case, the likelihood of returning to work at a given point in time) is constant over time for each categorical variable. For example, if men are twice as likely as women to return to work at 30 days, they should also be twice as likely to do so at 90 days, 180 days, and so forth. To verify that the proportional hazard assumption underlying the Cox model was met, we plotted the relative hazards over time for each categorical variable. All of the variables included met the proportional hazard assumption.

Because of the large number of claims in our sample, it was common for the coefficients to be highly significant but small in magnitude. Therefore, we imposed the additional criterion of "stability" in determining which results to give emphasis. In particular, we ran the survival analysis for each calendar year sample separately and compared the resulting coefficients across years. Results were classified as stable and significant if, in each of the 3 years, they (1) were statistically significant and (2) had the same sign and roughly the same magnitude (within a band of 50% of the average coefficient value).

Results

Figure 2 shows a survival curve (Kaplan-Meier) of the percentage of claims in the sample by duration of disability, unadjusted for covariates. The figure demonstrates that most disability du-

rations were short term; by 1 month, more than half of the employees had returned to work. Table 1 provides descriptive statistics for the variables used in the multivariate analysis. Of those filing claims, three quarters were men, only 17% were older than 45 years of age, and half were married at the time of the injury. The largest single category of injury was back sprains (34%), followed by other sprains and strains (20%). Six percent of the claims involved a hospitalization in the first 28 days after the date of injury. Slightly more than one third of the claims were filed in King County (Seattle metropolitan area), and 57% were filed by workers in small firms (fewer than 50 employees).

The mean disability durations for each category shown in Table 1 are included for purposes of illustration only. Censored claims were included in the calculations, and, since categories with claims of longer duration will also have more censored claims, the true differences in duration across categories (e.g., across age groups) are probably understated. Given this caveat, the unadjusted differences in duration across categories ranged from a few days to nearly 5 months in the case of injuries with and without hospitalization.

Table 2 presents results from a Cox proportional hazards regression that included all of the independent variables in Table 1 for the 1987 to 1989 combined sample. For each variable, the coefficient and standard error are shown, as well as the relative hazard (exponent of the coefficient) and its confidence interval. For example, women were only 84% as likely as men to return to work at any given point in time. As noted above, we distinguished between effects found to be stable and significant over the 3-year period from other effects that were either insignificant in some years or had large changes in sign and/or magnitude. The largest effects involved injuries requiring hospitalization (relative hazard = 0.48), carpal tunnel injuries (relative hazard = 0.55), and older workers (relative hazard = 0.67). The coefficients on the other stable and significant variables, while smaller in magnitude, indicate that workers with a longer duration of disability were more likely (1) to be women, older, and divorced; (2) to have dependents; and (3) to work in smaller firms. In addition, workers in counties with high unemployment rates and those in general construction were more likely to have claims of longer duration. Workers in wholesale trades were the least likely to have such claims.

TABLE 2—Survival Analysis: Cox Proportional Hazards Model, 1987–1989 Combined Sample

	Relative Hazard[a]	95% CI[b]
Stable and significant effects		
Gender		
Male	1.00	Reference
Female	0.85	0.82, 0.88
Age, y		
<30	1.00	Reference
30–44	0.78	0.76, 0.81
45+	0.67	0.64, 0.69
Family status		
Never married	1.00	Reference
Married	0.99	0.94, 1.04
Divorced	0.90	0.86, 0.94
Widowed	0.87	0.76, 1.00
Dependents		
No	1.00	Reference
Yes	0.88	0.85, 0.91
Type of injury		
Fracture	0.88	0.84, 0.92
Sprain, back/neck	0.79	0.77, 0.82
Sprain, other	0.87	0.84, 0.90
Carpal tunnel	0.55	0.50, 0.60
All other	1.00	Reference
Hospitalized		
No	1.00	Reference
Yes	0.48	0.46, 0.51
Firm size		
<50 employees	1.00	Reference
≥50 employees	1.08	1.05, 1.11
Unemployment rate, %		
<5	1.00	
5–7	0.99	0.95, 1.04
>7	0.88	0.82, 0.95

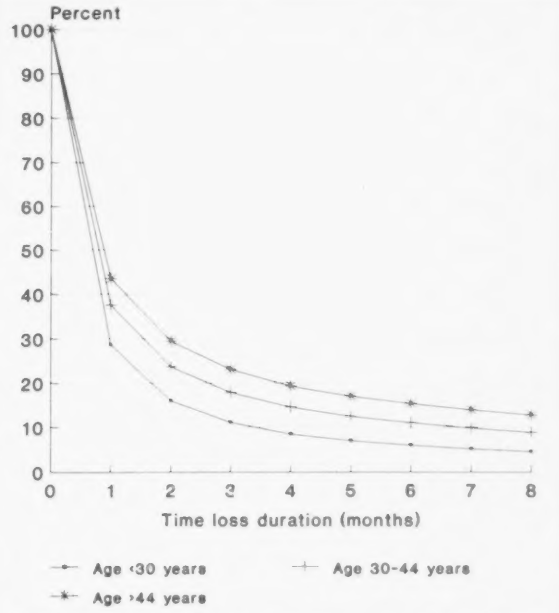
Some other results were noteworthy, even though they did not meet our criteria of stable and significant. Workers at firms participating in the retrospective rating program over the entire period had a significantly greater chance of returning to work (relative hazard = 1.14), although the effect of participation at the time of injury was insignificant. None of the benefit-rate variables were significant at $P = .01$, even in the combined sample of claims.

Figures 3 and 4 present results from the Cox regressions in an alternative, perhaps easier to interpret, form. The survival curves show the estimated disability survival probabilities (in percentage points), based on the model in Table 2, evaluated at durations of 1 to 8 months. The lower curve in each figure is the "reference" case, with values of 0 for all of the independent variables. In particular, the reference case model is male, under 30 years old, and unmarried; has no dependents; and has an injury other than sprains, fractures, and carpal tunnel syndrome that did not require hospitalization in the first 28 days after occurrence. The other curves show the effect of changing *only* the indicated variable. For example, the effect of being in the 30- to 44-year-old age range raises the estimated proportion of claims involving 6 months or more of disability from roughly 5% to 10%. The cumulative effect of changing several cat-

TABLE 2—Continued

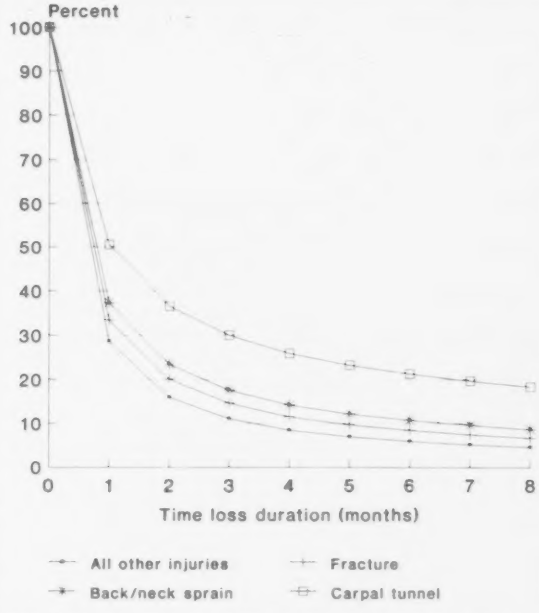
	Relative Hazard[a]	95% CI[b]
Other effects		
County		
King	1.00	Reference
Pierce	0.99	0.94, 1.05
Snohomish	0.91	0.86, 0.95
Spokane	0.84	0.79, 0.90
Other: east	0.97	0.92, 1.03
Other: west	1.02	0.96, 1.09
Year of injury		
1987	1.00	Reference
1988	0.99	0.96, 1.03
1989	1.03	0.99, 1.07
Industry		
Agricultural	0.88	0.83, 0.93
General construction	0.83	0.76, 0.90
Heavy construction	0.93	0.88, 0.97
Special trade construction	0.91	0.86, 0.95
Lumber products	0.91	0.86, 0.97
Trucking	1.09	1.04, 1.15
Wholesale trade	1.23	1.17, 1.29
Restaurant service	1.05	1.00, 1.10
Health services	1.02	0.95, 1.11
Auto dealers	0.90	0.84, 0.97
All other	1.00	Reference
Firm ownership		
Private	1.00	Reference
Government	1.06	1.00, 1.13
Retrospective rating program		
At time of injury		
No	1.00	Reference
Yes	1.00	0.95, 1.06
Participation over 4 years		
Not participating	1.00	Reference
In entire period	1.14	1.06, 1.22
Came in and stayed	1.08	1.02, 1.15
In and out of program	1.02	0.97, 1.07
Benefit rate, %		
<60	1.00	Reference
60–65	1.05	1.00, 1.10
>65	1.02	0.97, 1.06

[a]Conditional probability of returning to work relative to reference group.
[b]For relative hazard, confidence intervals (CI) containing 1.0 represent effects not significant at $P = .05$.



Note. Values were derived from "reference case" model (male, unmarried, not hospitalized, "other" injury type).

FIGURE 3—Percentage of workers receiving disability payments after the indicated number of months since injury, by age.



Note. Values were derived from "reference case" model (male, <30 years old, unmarried, not hospitalized).

FIGURE 4—Percentage of workers receiving disability payments after the indicated number of months since injury, by injury.

egories at once (not shown) can be quite large; for example, a married man more than 45 years old with dependents and an injury requiring hospitalization raises the proportion of claims involving at least 6 months of disability from 6% to 45.3%.

Discussion

Using a large random sample of all incident disability claims from the Washington State workers' compensation system, this investigation has shown that, after adjusting for initial hospitalization, factors that predicted longer duration of disability included older age, female gender, and a diagnosis of carpal tunnel syndrome or back/neck sprain. Other stable and significant predictors with lower magnitudes of effect included divorced marital status, firm size of fewer than 50 employees, higher county unemployment rates, and construction and agricultural work. Workers from wholesale trades were at decreased risk for longer term disability. Wage replacement had no significant effect on duration of disability.

This population-based study suggests that 17.5% of all initial disability claims involved at least 6 months of lost time, 12% involved 1 year of lost time, and 7.4% involved at least 2 years of lost time (see Figure 1). Only 12.1% of those claims in which at least 6 months of lost time accumulated involved injuries requiring hospitalization within 1 month of occurrence.

Before comparing our results with other studies, we should note that we defined disability as the loss of capacity to meet occupational demands as opposed to loss of personal or social functioning. The studies cited below share this focus on work-related impairment.

Many studies have focused on a specific type of injury and therefore have not been able to isolate the impact of injury type on duration. Studies including injury type have also found sprains and fractures to be associated with longer duration.[8] The most interesting injury-type finding here is the long duration of carpal tunnel claims, a result that was consistent across years and not apparently due to outliers. The carpal tunnel results should be viewed as preliminary given the small number of claims in this category (2%) and the relative crudeness of the method for identifying such claims (use of Z16.2 codes). Franklin et al.[24] found some misclassification even with more detailed medical information about the injury and more stringent criteria for classifying claims as carpal tunnel injuries.

Injury severity is another variable that has not been widely used in other studies. Our ability to characterize the nature and severity of the injury was limited by the quality of the medical billing database. We chose 1 month as the cutoff interval as a compromise between (1) ensuring that there was adequate time for the appropriate injury treatment to be chosen and (2) ensuring that the hospitalization shown in the billing record was primarily due to the particular work-related injury. Other studies that have used hospitalization as a proxy for severity[10] have also found it to be correlated with disabilities of longer duration. In our data, a hospitalization within the first 28 days was the strongest and most consistent correlate of longer term disability.

A number of worker and family characteristics, of which older age is the most important and consistent, have been found to influence duration of disability.[2,8–11] The age effect is due both to the reduced ability of older workers to recover from injuries and to the reduced likelihood of finding employment once they have recovered. Gender has not figured prominently in previous studies, partly because a much

smaller proportion of women file workers' compensation claims. In our analysis, women had disabilities of longer duration, as has been found in other studies in which gender has been included as a variable.[11] Divorced workers have also been found in other studies to have longer term disabilities.[3,4]

Several personal characteristics not available on the Washington database have been identified as correlates of disability duration, particularly indicators of socioeconomic status. Higher socioeconomic status is expected to be associated with fewer claims and shorter duration.[3,4,12] Manual labor occupations have also been associated with disabilities of longer duration when other factors have been controlled.[10,11] A number of studies have found that higher workers' compensation benefits result in longer duration.[5,7,9,10] We did not find an association between the wage replacement rate and duration; however, as noted above, the replacement rate was confounded with marital status and number of dependents. Also, because of the confounding problem, we did not pursue other, perhaps more accurate measures of the income replacement rate such as after-tax income.

Three British studies and one American study[9–11,14] have also reported finding a longer duration (or higher incidence) of disabilities in areas with high unemployment rates. This indicates that duration of disability may be sensitive to the prospects for employment in the immediate area. This result should be viewed with caution given the difficulties involved in estimating the effect of aggregate variables on individual-level behavior.[25]

Industry results are scarce in the literature,[6,9] and few consistent patterns have emerged. As noted above, the industry effects in our analysis were also not significant or consistent across years. In the combined sample, workers in construction and agriculture tended to have disabilities of longer duration, while disabilities in trade and services were of shorter duration.

Size was the single firm characteristic that was a stable and significant predictor of duration, with larger firms having shorter durations of disability. This result is consistent with other studies.[5,17] Drury[17] has suggested a number of reasons for smaller firms having more disability claims and longer durations of disability: the ability of larger firms to employ specialists in disability management, exemption of smaller firms from laws requiring disability benefits, and high turnover and less access to information about disability prevention in smaller firms. Another factor providing large firms more incentive to shorten claims is their greater experience ratings in standard premium calculations.[26] This predicted effect of experience rating is consistent with our finding that participants in the retrospective rating program (in which the experience rating is more complete) had claims of shorter duration. Finally, larger firms may have greater flexibility in allowing workers to return to modified jobs.

There are some qualifications and limitations to our results that should be noted. First, the data were from an administrative database and therefore subject to a certain amount of entry errors, miscoding, and misclassification.[24,27] Much of the information used was from the claim form filled out by the physician and coded by the state at the time of injury. Those fields required for billing and claims administration (e.g., marital status, company identifier, hospitalization records) are closely monitored; fields not directly related to billing the state and the firm may be subject to greater error (e.g., injury type). In general, the effect of coding and data-entry errors will be to attenuate the relationships examined in the survival analysis.

The results are obviously specific to Washington State, and caution should be exercised in extrapolating them to other states. Washington does have a relatively diverse set of industries, including resource-based (logging, agriculture), heavy and light manufacturing, and trade and services industries. However, a number of the largest employers—350 representing one third of covered workers—were self-insured and therefore excluded from the study. Firm size and the retrospective rating program were included in the analysis, so some of the effects of size and presence of disability prevention incentives similar to being self-insured were at least partially controlled. Other factors specific to Washington include the claims-handling procedure and system of vocational rehabilitation. A substantial revision of claims handling in early 1989 was intended to reduce the number of long-duration claims, although an effect of this change on disability duration was not found in our sample (Table 2).

A final limitation is that these independent variables measured worker, firm, and industry characteristics only at the time of injury. Other factors related to medical management and vocational rehabilitation were highly correlated with duration of disability; thus, their effects could not be estimated independently within our study design.

This study is subject to a number of limitations; however, its strength was that it (1) was population based, (2) allowed for a mean follow-up of nearly 3 years, and (3) adjusted for hospitalization within 28 days of injury (used as a proxy for injury severity). The principal findings suggest that older workers, women, and patients with carpal tunnel syndrome or low-back/neck sprains are all at a disadvantage in regard to the risk for longer term disability, even after adjustment for hospitalization. Greater disability prevention efforts targeted at these subgroups could have significant economic and public health effects. □

Acknowledgments

This work was carried out under a contract from the Washington State Department of Labor and Industries.

We wish to thank Stephanie Findley for her assistance in the project.

References

1. *Workers' Compensation Claim Characteristics.* New York, NY: National Council on Compensation Insurance; 1991.
2. White AWM. Low back pain in men receiving workman's compensation. *Can Med Assoc J.* 1969;101:61–67.
3. Bolton B. *Introduction to Rehabilitation Research.* Springfield, Ill: Thomas; 1974.
4. Volinn E, Van Koevering D, Loeser JD. Back sprain in industry: the role of socioeconomic factors in chronicity. *Spine.* 1991;16:542–548.
5. Ruser JW. Workers' compensation and occupational injuries and illnesses. *J Labor Economics.* 1987;9:325–350.
6. Chirikos TN. Accounting for the historical rise in work-disability prevalence. *Milbank Q.* 1986;64:271–301.
7. Butler RJ, Worrall JD. Work injury compensation and the duration of nonwork spells. *Econ J.* 1985;95:714–724.
8. Johnson WG, Ondrich J. The duration of post-injury absences from work. *Rev Economics Stat.* 1990;72:578–586.
9. Holmes P, Lynch M. An analysis of invalidity benefit claim durations for new male claimants in 1977/1978 and 1982/1983. *J Health Economics.* 1990;9:71–83.
10. Fenn P. Sickness duration, residual disability, and income replacement: an empirical analysis. *Econ J.* 1981;91:158–173.
11. Cornes P. The vocational rehabilitation index: a guide to accident victims' requirements for return-to-work assistance. *Int Disability Stud.* 1990;12:32–36.
12. MacKenzie EJ, Shapiro S, Smith RT, Siegel JH, Moody M, Pitt A. Factors influencing return to work following hospitalization for traumatic injury. *Am J Public Health.* 1987;77:329–334.
13. Johnson KL, Chan F, Questad K. Statistical prediction in proprietary rehabilita-

tion. *Rehabil Counseling Bull.* 1987;30(3): 130–135.
14. Volinn E, Lai D, McKinney S, Loeser JD. When back pain becomes disabling: a regional analysis. *Pain.* 1988;33:33–39.
15. Burton AK, Tillotson KM. Prediction of the clinical course of low-back trouble using multivariable models. *Spine.* 1991;16: 7–14.
16. Bigos SJ, Battie MC, Spengler DM, et al. A prospective study of work perceptions and psychosocial factors affecting the report of back injury. *Spine.* 1991;16:1–6.
17. Drury D. Disability management in small firms. *Rehabil Counseling Bull.* 1991;34: 243–256.
18. Breslow NE, Day NE. *Statistical Methods in Cancer Research II: Design and Analysis of Cohort Studies.* Lyon, France: International Agency for Research in Cancer; 1987. IARC Scientific Publication 82.
19. England P, Kilbourne B. *Occupational Measures from the Dictionary of Occupational Titles for 1980 Census Detailed Occupations.* Ann Arbor, Mich: Michigan Inter-university Consortium for Political and Social Research; 1988.
20. Stevens G, Hoisington E. Occupational prestige and the 1980 US labor force. *Soc Sci Res.* 1987;16:74–105.
21. Stevens G, Cho JH. Socioeconomic indexes and the new 1980 census occupational classification scheme. *Soc Sci Res.* 1985;14:142–168.
22. Cox DR. Regression models and life tables. *J Royal Stat Soc.* 1972;30:248–275.
23. Harrell FE. *The PHGLM Procedure. SUGI Supplemental Library User's Guide. Version 5 Edition.* Cary, NC: SAS Institute; 1986.
24. Franklin GM, Haug J, Heyer N, et al. Occupational carpal tunnel syndrome in Washington State, 1984–1988. *Am J Public Health.* 1991;81:741–746.
25. Moulton B. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Rev Economics Stat.* 1990; 72:334–338.
26. Russell LB. Safety incentives in workmens' compensation insurance. *J Hum Resources.* 1974;9:361–375.
27. Connell FA, Diehr P, Hart LG. The use of large databases in health care studies. *Annu Rev Public Health.* 1987;8:51–74.

Association for Worksite Health Promotion Sponsors April Conference

Health promotion and disease prevention are pivotal to successful health care reform both nationally and at every local level. A unique conference, "Health Promotion and Health Care Reform: Linking Values and Building Partnerships in the '90s," offers business and health professionals new strategies, modes, and insights on implementing effective worksite and community health promotion programs. The conference will be held April 17 through 19, 1994, at the San Francisco Airport Marriott Hotel, 1800 Bayshore Hwy, Burlingame, Calif.

Featured speakers will include Michael McGinnis, MD, deputy assistant secretary of health, US Office of Disease Prevention and Health Promotion; Ken Pelletier, MD, Stanford Center for Research in Disease Prevention; Susan Seidler, Washington Business Group on Health; Brick Lancaster, MPH, Centers for Disease Control and Prevention; Helen Schauffler, PhD, MSPH, UC Berkeley School of Public Health; and Carson Beadle, The Health Project/William M. Cercer, Inc.

Sponsored by the Association for Worksite Health Promotion (Region 7), an international nonprofit professional organization, the conference is designed for leaders in human resources, benefits, health promotion, and health insurance, as well as occupational health nurses, physicians, and other clinical practitioners. Cosponsors include Lifeguard, Inc., Parke-Davis, Marion Merrell Dow, Merck, Lederle, Bay Sport, Xerox, Longs Drug Stores, Health Dimensions, and Richard P. Carr Physical Therapy.

For registration and conference information, please call (415) 985-2698.

Circumcision and Sexually Transmitted Diseases

Linda S. Cook, MS, Laura A. Koutsky, PhD, and King K. Holmes, MD, PhD

ABSTRACT

Objectives. New evidence linking lack of circumcision with sexually transmitted human immunodeficiency virus revives concerns about circumcision and other sexually transmitted diseases. This study was undertaken to assess the relationship between circumcision and syphilis, gonorrhea, chlamydial infection, genital herpes, nongonococcal urethritis, and exophytic genital warts.

Methods. A cross-sectional study of 2776 heterosexual men attending a sexually transmitted disease clinic in 1988 was used to investigate the relationship between circumcision and sexually transmitted diseases. Subjects with specific sexually transmitted diseases and those without such diseases were compared after adjustment for age, race, zip code of residence, other sexually transmitted diseases, and number of sexual partners.

Results. A positive relationship was observed between uncircumcised status and both syphilis and gonorrhea. A negative relationship was found between warts and lack of circumcision. No apparent relationship was noted between uncircumcised status and genital herpes, chlamydial infection, or nongonococcal urethritis.

Conclusions. Uncircumcised men were more likely than circumcised men to have syphilis and gonorrhea and were less likely to have visible warts. (*Am J Public Health.* 1994;84:197–201)

Introduction

The possible role of male circumcision in the acquisition of sexually transmitted diseases has taken on new significance with recent evidence that human immunodeficiency virus (HIV) infection in men is associated with uncircumcised status.[1–3] Sexually transmitted diseases previously linked with uncircumcised status include chancroid, syphilis, gonorrhea, and genital herpes infection[1,4–7]; however, most of the reported associations between these diseases and uncircumcised status have not been adequately adjusted for potentially confounding factors. The present cross-sectional study was undertaken to examine the relationship between circumcision status and genital infection by *Treponema pallidum*, *Nisseria gonnorhoeae*, and *Chlamydia trachomatis*, in addition to clinical evidence of nongonococcal urethritis, genital herpes, and exophytic genital warts, with adjustment for potentially confounding factors.

Methods

Study Population

The study population included men attending the Seattle–King County Department of Public Health Sexually Transmitted Disease Clinic at Harborview Medical Center between January and December 1988. The clinic, which provides comprehensive services related to sexually transmitted diseases such as diagnosis, treatment, education, and risk reduction counseling, served 80 to 120 patients per day in 1988. The exact number of unique men seen in the clinic in 1988 is unknown; the Department of Public Health tallies new problem visits, to which an individual can contribute multiple times. Among all male visits to the clinic in 1988, 8034 new problem visits were recorded.

Study subjects were identified in 1990 through patient medical charts that had been archived according to the year of visit. Charts are placed into the archives after approximately 1 year if the patient has not returned to the clinic in the interim. Charts of men who returned to the clinic in 1989 were not included in the study. According to statistics compiled by the clinic, the study population closely matched that of all men seen for a new problem in 1988 in regard to cases of syphilis (0.7% vs 0.8%), gonorrhea (8.1% vs 10.0%), warts (14.5% vs 14.2%), chlamydial infection (6.8% vs 6.9%), nongonococcal urethritis (29.0% vs 25.4%), and genital herpes (4.7% vs 5.0%). Thus, the study population was considered a good representative sample. Of the 4137 subjects identified, 897 did not meet the study inclusion criteria (380 who presented for HIV screening only and did not have a genital examination, 306 who were homosexual or bisexual, and 211 who did not reside in King County). Of the 3240 eligible study subjects, 464 did not have circumcision status recorded and were excluded from the analysis. The resulting study population consisted of 2776 heterosexual men.

All data used in this study were abstracted in 1990 from standardized clinic records of men who attended the clinic in

The authors are with the Department of Epidemiology, School of Public Health, University of Washington, Seattle. Laura A. Koutsky and King K. Holmes are also with the Center for AIDS and STDs, School of Medicine, University of Washington.

Requests for reprints should be sent to Laura A. Koutsky, PhD, Center for AIDS and STDs, 1001 Broadway, Seattle, WA 98122.

This paper was accepted May 11, 1993.

1988. These records included information concerning demographic characteristics of the patient, reason(s) for visit, symptoms, sexual history (new sexual partner in the last month, number of sexual partners in the last month, and days since last sexual exposure), prior sexually transmitted disease history, and results of all physical and laboratory examinations.

All consenting patients seen for a new problem underwent genital examination and had urethral swabs taken for a Gram-stained smear and for cultures for *N. gonorrhoeae* and *C. trachomatis*. Blood for serologic testing of syphilis (reagin antibody) was obtained from all consenting patients, and dark-field examination was performed on suspected syphilitic lesions.

Diagnoses of warts and genital herpes were presumptively based on the clinical appearance of lesions; cultures for herpes simplex virus were obtained when considered necessary by the clinician. Current syphilis was determined by a positive dark-field test from a suspicious lesion or by a positive rapid plasma reagin (RPR) or Venereal Disease Research Laboratory (VDRL) titer (titer above 1:2) with confirmatory T. pallidum hemagglutination assay (TPHA) serology. This definition included primary, secondary, and some early latent syphilis cases; however, all cases were newly diagnosed.

The diagnosis of nongonococcal urethritis was based on the presence of at least two of the following criteria: subject-reported urethral discharge or dysuria, presence of a purulent urethral discharge on exam, or a urethral Gram-stained smear that showed at least five polymorphonuclear leukocytes per 1000X oil immersion field for at least three fields. The combination of the first two criteria was used only if the *N. gonorrhoeae* culture was negative. A subset of subjects (n = 278) had elevated polymorphonuclear leukocytes on urethral smear; these subjects were given a clinical diagnosis of nongonococcal urethritis and treated for this condition but were not recorded as having an abnormal urethral discharge. Addition of these subjects to those in the more stringently defined category of nongonococcal urethritis did not alter the study results, so they were combined in the final analysis.

For this study, the principal sexually transmitted diseases evaluated were laboratory-confirmed *N. gonorrhoeae, C. trachomatis*, and *T. pallidum* infection and clinical evidence of nongonococcal urethritis, genital warts, or genital herpes. The "other sexually transmitted disease" category comprised subjects with a variety of diagnoses, including pediculosis pubis (n = 81), scabies (n = 33), and molluscum contagiosum (n = 69). The reference group was composed of subjects who were not diagnosed with any of the above sexually transmitted diseases or conditions. Circumcision status was recorded by the clinician as part of the routine examination. Place of residence in King County was categorized by zip code into six Seattle areas and one non-Seattle area, as explained in detail elsewhere.[8]

Statistical Analysis

The initial 1988 clinic visit of each subject was used for analysis. The prevalence of each of the principal study diseases was determined on the basis of circumcision status. In order to allow direct sexually transmitted disease prevalence comparisons between circumcised and uncircumcised men, prevalences were standardized to the combined age and race distribution of the entire study population.

Characteristics of those with sexually transmitted diseases and those in the reference group without such diseases were compared. Odds ratios (ORs) were used as the measure of association to assess the relationship between circumcision status and each disease category. Logistic regression[9,10] was used to calculate adjusted odds ratios and the associated 95% confidence intervals (CIs). Since subjects could have multiple sexually transmitted disease diagnoses, all variables were placed in a single model to allow for the control of multiple diseases. Therefore, adjusted odds ratios reflect an association between a particular sexually transmitted disease and circumcision status that is not influenced by the presence of other sexually transmitted diseases. All analyses were conducted with SAS.

Results

Prevalences of Sexually Transmitted Diseases Under Study

Age- and race-standardized prevalence figures revealed that, in this clinic population, the prevalence of current syphilis was higher in uncircumcised men than in circumcised men (2.5% vs 0.6%; $P < .001$), as was the prevalence of gonorrhea (12.7% vs 8.1%; $P < .01$). Conversely, the prevalence of exophytic warts was higher in circumcised men (17.6% vs 11.0%; $P < .001$). No significant differences between uncircumcised and circumcised men were observed for prevalences of nongonococcal urethritis, (30.2% vs 32.7%), genital herpes (8.8% vs 8.7%), or urethral infection with *C. trachomatis* (7.2% vs 6.4%).

Subject Characteristics Related to Current Diagnosis

The study population, composed of 540 uncircumcised men (19.5%) and 2236 circumcised men (80.5%), is described in Table 1 according to sexually transmitted disease status. In comparison with the reference group, those with gonorrhea, chlamydial infections, and genital warts were younger ($P < .05$); subjects with a diagnosis of current syphilis, gonorrhea, chlamydial infection, or nongonococcal urethritis were more likely to be African American, whereas those with exophytic warts were more likely to be White. Subjects with current diagnoses of gonorrhea, chlamydial infection, and nongonococcal urethritis more often reported a new sexual partner in the previous month and two or more partners in the previous month than those in the reference group. Subjects diagnosed with any of the principal study sexually transmitted diseases were more likely than those with no disease diagnosis to present at the clinic with symptoms and less likely to present for screening.

Association of Uncircumcised Status with Current Disease Diagnosis

Crude and adjusted odds ratios for the association of uncircumcised status with the study diseases are presented in Table 2. Adjusted odds ratios were estimated by fitting a multiple logistic regression model adjusted for age, race, number of sexual partners in the previous month, place of residence, and the other sexually transmitted diseases. No adjustment was made for a new partner in the last month, days since last sexual exposure, marital status, and reason for visit because of their negligible effects on the estimated odds ratios. No adjustment was made for a prior history of sexually transmitted diseases because any influence of circumcision status was assumed to be the same for a prior diagnosis as for the current diagnosis, and adjustment would have controlled for the associations the present study sought to quantify. In the adjusted analysis, an elevated odds ratio was noted for the association of uncircumcised status with current syphilis (OR = 4.0, 95% CI = 1.9, 8.4) and gonorrhea (OR = 1.6, 95% CI = 1.2, 2.2). A negative association was noted for warts and uncircum-

TABLE 1—Subject Characteristics (%), According to Sexually Transmitted Disease Status

	Reference Group[a] (n = 985)	Syphilis (n = 33)	Gonorrhea (n = 262)	Chlamydial Infection (n = 181)	Genital Herpes (n = 254)	Nongonococcal Urethritis (n = 882)	Genital Warts (n = 463)	Other Sexually Transmitted Diseases (n = 180)
Age, y								
11–19	8	6	13*	19***	6	10	7	17***
20–24	20	27	28*	30**	21	23	27*	39***
25–29	24	27	23	22	26	26	29*	22
30–34	17	12	17	13	21	15	18	7***
35+	27	27	17**	12***	21	21	14***	11***
Unknown	6	...	3	4	6	5	5	4
Race/ethnicity								
White	61	9***	21***	58	65	57***	73***	62
African American	23	61**	60***	31***	21	33***	19	25
Other	13	27	15	9	10	9**	5***	10
Unknown	3	3	4	2	4	1	3	3
Marital status								
Single	84	82	91**	93**	85	88***	86	89*
Married	11	12	6**	4**	10	8***	12	4**
Unknown	4	6	3	2	4	4	3	6
No. of sexual partners in previous month								
0	14	18	1***	4***	10	6***	13	15
1	60	58	55*	61	65	63	65	57
2+	18	21	41***	30***	21	27***	17	23
Unknown	7	3	3*	4	3*	3*	5	4
New sexual partner in previous month	31	33	63***	50***	26	43***	21**	34*
No. of days since last sexual exposure								
0–7	50	33	69***	50	59	54	50	48
8–14	13	24*	19**	27**	14	23***	16*	19*
15+	30	36	9***	22	20**	20***	26	28
Reason for visit								
Symptoms	51	55	89***	74***	86***	71***	70***	80***
Contact with:								
Gonorrhea	5	3	8	7	1**	3	2**	3
Syphilis	0.8	33**	...	...	...	0.2	0.1	...
Chlamydia	6	3	1***	17**	2***	13	4**	4*
Other sexually transmitted disease	14	3	2***	8**	14	12**	12	9*
Screen	28	3	5***	8***	9***	13***	16***	11***
Consultation	5	12	1**	3	3	2*	5	6

[a]Group with no sexually transmitted diseases.
*.01 < *P* < .05.
**.001 < *P* < .01.
***$P < .001$ (in comparison with the reference group).

cised status (OR = 0.7, 95% CI = 0.5, 0.9). Uncircumcised status was not associated with nongonococcal urethritis, genital herpes, or urethral infection with *C. trachomatis*.

Discussion

The results of this study support the hypothesis that circumcision status is related to some, but not all, sexually transmitted diseases. The positive findings associating syphilis and gonococcal infections with uncircumcised status in the present study were similar to those previously reported by the Public Health Department Special Treatment Clinic in Perth, Australia.[6] Age-adjusted results from the Perth study indicated a positive relationship with both syphilis (OR = 5.4, 95% CI = 1.4, 21.1) and gonorrhea (OR = 2.3, 95% CI = 1.5, 3.7). Contrary to results in the present study, the Perth study also found a positive relationship between uncircumcised status and herpes and warts. Discrepancies between the two studies could be due to the present study's finer adjustment for confounding variables and restriction to heterosexual men.

Other studies have also investigated specific relationships between certain sexually transmitted diseases and uncircumcised status. In a study conducted at the Whitechapel Clinic in London, a comparison of men with culture-confirmed herpes and controls revealed a higher proportion of men with herpes being uncircumcised.[5] Interpretation of this finding is limited since the control group consisted of men with a number of other sexually transmitted disease diagnoses that each could have had a specific relationship with circumcision status. A study of active duty army men found that uncircumcised status was protective for nongonococcal urethritis

TABLE 2—Odds Ratios Associating Study Sexually Transmitted Diseases with Uncircumcised Status

	Crude Distribution, No.		Crude Odds Ratio	Adjusted Odds Ratio[a]	95% Confidence Interval
	Uncircumcised	Circumcised			
Reference group	198	787	1.0	1.0	...
Syphilis	20	13	6.1	4.0	1.9, 8.4
Gonorrhea	87	175	2.0	1.6	1.2, 2.2
Chlamydial infection	34	147	0.9	1.0	0.7, 1.6
Nongonococcal urethritis	161	721	0.9	1.0	0.8, 1.3
Genital herpes	49	205	1.0	1.2	0.8, 1.7
Genital warts	51	412	0.5	0.7	0.5, 0.9
Other sexually transmitted diseases	26	154	0.7	0.8	0.5, 1.2

[a]Adjusted for age group (13–19, 20–24, 25–29, 30–34, 35+), race/ethnicity (White, African American, other), number of sexual partners in the last month (0, 1, 2+), place of residence (six Seattle areas and one non-Seattle area defined by zip codes), and the other sexually transmitted diseases.

but not related to gonococcal infection.[11] Although that study adjusted for important confounders, direct comparisons with the present study are difficult since the former included subjects who apparently contributed to the case groups several times. In the final analysis, there were approximately five times more men in both the gonorrhea and nongonococcal urethritis groups than in the control (no sexually transmitted disease) group. In addition, the ratio of nongonococcal urethritis cases to gonorrhea cases was much lower than that typically observed in sexually transmitted disease clinic populations. Finally, a study of Canadian army men reported a higher proportion of uncircumcised active duty men with warts and syphilis compared with the proportion of uncircumcised men among new recruits.[12] Although suggestive, the results could easily be biased as a result of the comparison of two distinctly different groups of men.

It was not possible to evaluate chancroid because it was not found in this study population. It was also not possible to evaluate HIV infection as a result of its very low prevalence (estimated to be 0.6% among heterosexual men seen in the clinic from June 1988 through December 1990). Therefore, no comparisons can be made with the many studies linking uncircumcised status with chancroid, including recent studies from East Africa that have linked uncircumcised status with both chancroid and HIV seroconversion.[1-3]

Mechanisms for differential risks of sexually transmitted diseases based on circumcision status remain undefined. Possible theories include the following: (1) trauma of the intact foreskin during sexual intercourse produces microscopic lesions that increase susceptibility; (2) the environment under the foreskin could extend the survival of infectious agents and thus prolong exposure to viable organisms; (3) the thinner, less cornified epithelium on the glans and preputial sac of an uncircumcised male may provide less of a physical barrier to microbial invasion; and (4) nonspecific balanitis, not uncommon in uncircumcised men, may predispose to certain sexually transmitted diseases.

The finding that circumcision status is associated with some but not all sexually transmitted diseases may be related to the differing pathogenesis of these diseases. For syphilis, entry by *T. pallidum* is thought to occur through microscopic abrasions and not through undamaged, keratinized epithelium.[4] In uncircumcised men, trauma of the foreskin or the thin epidermis of the glans during sexual activity can open a passageway of infection for treponemes. The thicker, cornified epithelium on the glans of a circumcised male may be less susceptible to abrasions and, therefore, to entry by *T. pallidum*. For gonorrhea, the inner lining of the foreskin[13] might create an environment conducive to the survival or growth of *N. gonorrhoeae* and, perhaps, to entry into the urethra and enhanced infectivity.

A limitation of the present study arises from the available information on genital herpes and genital warts. For genital herpes diagnosis was based on the presence of clinical lesions, with or without confirmatory cultures, and subclinical infections could have been undetected. If uncircumcised subjects were more likely to have inapparent infections, this study could have underestimated the true association of uncircumcised status with genital herpes. The finding that genital warts were less likely to be diagnosed in uncircumcised men was of interest. Further information on the anatomic distribution of penile warts is actually the relevant information for assessing the influence of circumcision status on wart lesions and will be addressed in a subsequent paper.

A possible concern is that the study subjects presumably came to the clinic as a result of a problem related to a sexually transmitted disease and that, therefore, an elevated prevalence of one disease would predict a lower prevalence of some other disease. In this cross-sectional population, subjects were not selected on the basis of the presence of symptoms related to, or a diagnosis of, a sexually transmitted disease. Thirty-five percent of the study population did not have a sexually transmitted disease diagnosis. Furthermore, a subject could be diagnosed with multiple diseases, and, because testing for infection with *N. gonorrhoeae* and *C. trachomatis* and for syphilis was routine in all patients, diagnoses were not mutually exclusive. Since neither a sexually transmitted disease diagnosis nor a single diagnosis were obligatory for inclusion in the study, the prevalence of one disease should not necessarily have influenced the prevalence of another disease.

Because the study population was composed of men presenting at a public clinic, we were unable to control for certain self-selection factors in the analysis. The frequency of exposure to sexually transmitted diseases, health-seeking behavior, and perception of disease could all influence whether or not a person presents at a clinic. Therefore, the results can be generalized only to men who, given the possibility of a sexually transmitted disease, would present at a clinic. Men who do not seek care or who seek care from clinicians in private practice are likely to be different from the men who present at clinics such as the one described here.

In addition, the study population represented a sample of all heterosexual male visits to the clinic in 1988. It is possible that the proportion of uncircumcised and circumcised men in the study population differed from the total 1988 clinic population. Of particular concern are the men who returned to the clinic in 1989 and thus were not included in the study population as a result of their charts not being archived. This potential selection bias cannot be evaluated directly because the circumcision status of all men seen at the clinic in 1988 or 1989 was not tallied in health department statistics. However, the health department does report the age and race/ethnicity distributions for all male heterosexual clinic visits. Since these two characteristics were the strong-

est predictors of circumcision status in the study population, we compared the age and race/ethnicity distributions of the study population, all clinic visits of male heterosexuals in 1988, and all clinic visits of male heterosexuals in 1989. No significant differences were observed between the three groups, indicating a similar composition in terms of age, race/ethnicity, and, presumably, circumcision status. This implied that the source of the study population, the archived clinic charts, was a reasonable reflection of all heterosexual male clinic visits in 1988. Furthermore, the age and race/ethnicity distributions did not fluctuate from year to year, and, presumably, neither did the circumcision status of heterosexual men attending the clinic. Therefore, although this potential bias cannot be dismissed, we believe it is unlikely that it had a major impact on the study results.

As with all cross-sectional studies, it is difficult to distinguish newly acquired cases from more established cases, especially for sexually transmitted diseases such as warts and herpes. However, all of the study cases were classified as new problem visits at the clinic. The reason that some individuals delayed in seeking medical care is unknown but could be due to variations in disease perception, as well as in behavior with regard to seeking health care. As a result of this limitation, we addressed only the issue of who was or was not diagnosed with a sexually transmitted disease at their initial 1988 clinic visit, not acquisition per se.

A unique feature of circumcision is that it occurs almost exclusively in infancy or early childhood in the United States. A very small percentage of men require or request the procedure as adults. Therefore, we can safely assume that in almost all of the men attending the clinic, circumcision status was determined prior to the occurrence of the disease. Thus, although a cross-sectional design was used to investigate the association between circumcision and sexually transmitted diseases, the temporal sequence of events was not in question.

The determination of circumcision status is not always straightforward because of incomplete circumcisions and foreskins that naturally retract to some degree. Information from a larger database of patients seen in the clinic from 1980 to 1989 was used to compare the recorded circumcision statuses for any patient with multiple visits. Of the 5666 patients in the database, 2266 (40%) were seen more than once. Only 13 subjects had conflicting circumcision statuses recorded. Therefore, we are confident that the clinicians consistently identified and recorded circumcision status in the vast majority of subjects.

Benefits and disadvantages of male circumcision continue to be controversial, although recent research has attempted to address pertinent health issue associated with circumcision status.[14,15] The results of this study do not show a definitive benefit of circumcision; although uncircumcised men were more likely to have syphilis and gonorrhea, circumcised men were more likely to have genital warts and equally likely to have nongonococcal urethritis and genital herpetic lesions. □

Acknowledgments

This work was supported by National Institutes of Health grant AI27757.

The research described in this paper was presented, in poster form, at the annual meeting of the Society for Epidemiologic Research, June 1991, Buffalo, NY.

References

1. Simonsen JN, Cameron DW, Gakinya MN, et al. Human immunodeficiency virus infection among men with sexually transmitted diseases. *N Engl J Med.* 1988;319: 274–278.
2. Cameron DW, D'Costa LJ, Maitha GM, et al. Female to male transmission of human immunodeficiency virus type I: Risk factors for seroconversion in men. *Lancet.* 1989;2:403–407.
3. Hira SK, Kamanga J, Macuacua R, Mwansa N, Cruess DF, Perine PL. Genital ulcers and male circumcision as risk factors for acquiring HIV-1 in Zambia. *J Infect Dis.* 1990;161:584–585.
4. Holmes KK, Mardh P, Sparling PF, Wiesner PJ, eds. *Sexually Transmitted Diseases.* New York, NY: McGraw-Hill; 1990.
5. Taylor PK, Rodin P. Herpes genitalis and circumcision. *Br J Venereal Dis.* 1975;51: 274–277.
6. Parker SW, Stewart AJ, Wren MN, Gollow MN, Straton AY. Circumcision and sexually transmissible disease. *Med J Aust.* 1983;2:288–290.
7. Thirumoorthy T, Sng EH, Doraisingham S, Ling AE, Lim KB, Lee CT. Purulent-penile ulcers of patients in Singapore. *Genitourin Med.* 1986;62:252–255.
8. *Washington State/Seattle-King County HIV/AIDS Quarterly Epidemiology Report.* Seattle, Wash: Washington State Health Department; second quarter, 1991.
9. Breslow N, Powers W. Are there two logistic regressions for retrospective studies? *Biometrics.* 1978;34:100–105.
10. Breslow NE, Day NE. *Statistical Methods in Cancer Research. The Analysis of Case-Control Studies.* Lyon, France: International Agency for Research on Cancer; 1980.
11. Smith GL, Greenup R, Takafuji ET. Circumcision as a risk factor for urethritis in racial groups. *Am J Public Health.* 1987; 77:452–454.
12. Wilson RA. Circumcision and venereal disease. *Can Med Assoc J.* 1947;56:54–56.
13. Ross MH, Reith EJ. *Histology, a Text and Atlas.* New York, NY: Harper & Row; 1985.
14. Wiswell TE. Routine neonatal circumcision: a reappraisal. *Am Fam Physician.* 1990;41:859–863.
15. American Academy of Pediatrics. Report of the Task Force on Circumcision. *Pediatrics.* 1989;84:388–391.

Employment Status and Depressive Symptoms in Middle-Aged Women: A Longitudinal Investigation

Joyce T. Bromberger, PhD, and Karen A. Matthews, PhD

ABSTRACT

Objectives. This study investigated the psychological effects of paid employment and nonemployment, cross-sectionally and longitudinally, in 541 middle-aged women.

Methods. Women were seen for an extensive baseline evaluation and were reevaluated in an identical manner 3 years later. They were divided into four "synthetic" cohorts, based upon employment status at these two assessments.

Results. At baseline, nonemployed women reported higher levels of depressive symptoms than employed women, with nonemployed women with less education, low support from family and friends, or low marital satisfaction the most symptomatic. Symptoms decreased in women who were nonemployed initially but employed at the later exam; in all other groups, symptoms increased. Unlike the cross-sectional findings, however, the longitudinal findings showed that women who were nonemployed at baseline and had low levels of education, social support, or marital satisfaction did not benefit more from subsequent employment than did their counterparts who had high levels of these factors.

Conclusions. New paid work may have a positive effect on mood for middle-aged women. However, given the contradictions between our cross-sectional and longitudinal results, caution should be exercised in drawing conclusions from cross-sectional studies that likely confound employment status, symptoms, and moderating variables. (*Am J Public Health.* 1994;84:202–206)

Introduction

In the last 30 years, the number of women participating in the labor force has increased dramatically.[1,2] According to recent census information, 57% of women are currently in the civilian labor force, compared with 38% in 1960.[2] Whether there are positive or negative consequences concomitant with this trend is of central concern to public health. According to the "scarcity" hypothesis, employment adds another major role for women (besides homemaker and mother) that inevitably leads to role strain, overload, and psychological problems.[3] The alternative "enhancement" hypothesis argues that the occupancy of multiple roles enhances feelings of security and self-worth, and increases sources of gratification and social support[3]; thus, employment should improve women's mental health.

Cross-sectional studies of women's employment have shown either a positive or a null effect for women in general[4–7] and differential effects for certain subgroups of women: employment has been found to be more beneficial for women of low socioeconomic status,[4,8,9] for women whose husbands share responsibility for household chores or child care,[10–12] for married women in unsatisfying marriages,[4,13] and for women who report high job satisfaction compared with their appropriate counterparts. The correlational nature of cross-sectional research, however, precludes an assessment of the direction of effects in these observed relationships.

Two longitudinal investigations are relevant here. Among women aged 40 to 45 who were (1) unmarried or (2) married and Black, those who were out of the work force at the start of the study reported a greater increase in health problems over 5 years than those who were not.[14] Among young army wives, those who were not employed when the study began but were employed 6 months later had the largest overall increase in general psychological well-being when compared with women who were either employed or not employed at both times, or with women who were employed initially but not later.[15] Those who were employed at both times reported a modest decrease in well-being.

The present longitudinal study evaluated the relationship between change in employment status and change in depressive symptomatology in middle-aged women. Based on the "enhancement" theory, we anticipated that paid employment may be particularly important for the psychological and emotional well-being of women in midlife because it offers them new options in the face of the biological and social changes known to accompany middle age.[16] We specifically hypothesized that employment would be emotionally beneficial because it increases financial resources and provides additional sources of personal worth, gratification, and social support. Furthermore, the greatest psychological gains from employment would be experienced by women with few initial resources (i.e., low education and low social supports), by married women with unsatisfying marriages, and by women for whom paid work would not function as an overload (i.e., women with low household/child responsibilities). At the same time, working women who had positive job experiences (e.g., high job satisfaction) at baseline would be most negatively affected by subsequent nonemployment.

The authors are with the University of Pittsburgh School of Medicine, Pittsburgh, Pa.

Requests for reprints should be sent to Joyce Bromberger, PhD, Department of Psychiatry, University of Pittsburgh, 3811 O'Hara St, Pittsburgh, PA 15213.

This paper was accepted September 20, 1993.

Finally, the results from this longitudinal study were expected to replicate the results of our initial cross-sectional analyses.

Methods

Subjects

Between 1983 and 1985, 541 middle-aged women were recruited to participate in a prospective investigation of the effects of menopause on behavioral and biological cardiovascular risk factors.[17] Randomly selected women aged 42 to 50 who lived within selected zip codes in Allegheny County (including Pittsburgh, Pa) and had driver's licenses were sent a letter of invitation. Of the 2405 women subsequently contacted by telephone, 901 were eligible for inclusion, and of these, 541 (60%) agreed to participate. To be eligible for the study, participants had to have had menstrual bleeding within the prior 3 months; no surgical menopause; diastolic blood pressure below 100 mmHg; and no medications known to influence the biological risk factors under study, including psychotropic agents. Participants were better educated than nonparticipants.

At study entry, 411 of the women were employed and 129 were nonemployed; 3 years later, 524 of these women (97%) participated in a follow-up examination. One woman was excluded at baseline because she did not complete the Beck Depression Inventory. Based upon employment status at the baseline and 3-year assessments, women were divided into four "synthetic" cohorts: (1) always employed (employed at both visits), (2) always nonemployed (nonemployed at both visits), (3) newly nonemployed (employed at baseline and nonemployed at 3-year exam), and (4) newly employed (nonemployed at baseline and employed at 3-year exam).[18] Among those always employed, 96% were continuously employed throughout the study period; among those always nonemployed, 71% were continuously nonemployed.

The sociodemographic characteristics of these groups are shown in Table 1. Those who were always employed were more likely to be unmarried than the others and to be better educated than those who were always nonemployed; the latter group had more married women and fewer women with advanced degrees than all the other groups. Marital status and educational attainment were included as covariates in the analyses reported below. Perimenopausal and postmenopausal women were evenly distributed among the employment groups.

TABLE 1—Sociodemographic Characteristics of Four Employment Groups (n = 524)

	Always Employed (n = 348; 66.4%)		Always Nonemployed (n = 84; 16.0%)		Newly Nonemployed (n = 49; 9.4%)		Newly Employed (n = 43; 8.2%)		
	No.	%	No.	%	No.	%	No.	%	$P(\chi^2)$
Education									
High school/less	86	24.7	35	41.7	18	36.7	12	27.9	
Some college	74	21.3	18	21.4	11	22.4	10	23.3	
College degree	83	23.9	26	31.0	10	20.4	8	18.6	
Advanced degree	105	30.2	5	6.0	10	20.4	13	30.2	.001
Marital status (baseline)									
Married	229	65.8	80	95.2	40	81.6	34	79.1	
Unmarried	119	34.2	4	4.8	9	18.4	9	20.9	.000
Race									
Black	31	8.9	8	9.5	4	8.2	3	7.0	
White	317	91.1	76	90.5	45	91.8	40	93.0	.967
Age (3rd year)									
≤50 y	144	41.4	32	38.1	17	34.7	19	44.2	
>50 y	204	58.6	52	61.9	32	65.3	24	55.8	.743
Menopausal status									
Premenopausal	207	59.5	46	54.8	25	51.0	26	60.5	
Perimenopausal	62	17.8	22	26.2	10	20.4	5	11.6	
Postmenopausal	79	22.7	16	19.0	14	28.6	12	28.9	.394

Procedures

After interviews were conducted to determine eligibility and verify blood pressure levels, all participants were seen at a university clinic for a baseline assessment and, approximately 3 years later, for another evaluation. As part of both assessments, the women completed—in addition to measures of biological characteristics and health behavior—a self-report inventory providing data on employment status, job characteristics, depressive symptomatology, and various other psychosocial factors. Following the 3-year exam, the women were requested by letter to complete an employment history questionnaire.

Measures

Outcome variable. Depressive symptoms experienced during the 2 weeks prior to the clinic visit were measured by the Beck Depression Inventory. This 21-item self-report inventory rates the extent to which a symptom has been present; rates range from 0 (not at all/a little) to 3 (a lot/almost all the time). It has been shown to have good reliability and validity.[19] In the present sample, internal reliability was very good ($\alpha = .85$). Beck scores were subjected to a logarithmic transformation because of its skewed distribution in the present sample.

Independent variables. For conceptual and analytic consistency, all moderators were treated as categorical variables, with median splits used unless otherwise indicated. Less education refers to some college or less, and more education refers to a college or advanced degree.

Employment variables consisted of four job characteristics: job satisfaction, use of skills, job future ambiguity, and boss support. The first three were derived from a factor analysis of a 15-item questionnaire taken from an inventory developed by Caplan and associates[20] and modified by Cottington.[21] All factors demonstrated good internal reliability (αs = .69 to .82). Boss support was assessed by a 7-item yes/no scale that was developed in the Framingham Heart Study and yielded an alpha coefficient of .82 in that study.[22] In the present sample, the alpha coefficient was .70.

Married women indicated their level of satisfaction with seven aspects of their marital relationship ranging from 0 (not at all satisfied) to 3 (very satisfied). A factor analysis of these items yielded only one factor, which explained 60.1% of the variance and had very good internal reliability ($\alpha = .87$). Women then indicated how the responsibilities for seven areas of household and family life were divided between themselves and their husbands; such responsibilities were deemed (1) largely their husband's, (2) about equal, or (3) largely their own. A factor analysis of

TABLE 2—Baseline Mean (Log Transformed) Beck Depression Inventory Scores, Adjusted for Education and Marital Status,[a] According to Baseline Psychosocial Characteristics

	Employed	Nonemployed	Contrast *F* from Analysis of Covariance
Education			
Low	.60 (4.6)	.71 (6.4)	
High	.51 (3.6)	.59 (4.6)	5.48
Perceived social support			
Low	.67 (5.3)	.80 (8.1)	
High	.50 (3.5)	.58 (4.2)	16.23
Family support			
Low	.62 (4.9)	.82 (9.0)	
High	.54 (3.9)	.61 (4.7)	10.72
Friend support			
Low	.62 (4.7)	.80 (7.2)	
High	.53 (3.9)	.60 (5.0)	10.27
Marital satisfaction[b]			
Low	.62 (4.8)	.76 (7.2)	
High	.44 (2.7)	.46 (3.0)	27.47

Note. All *P*s <.05. The numbers in parentheses are the untransformed mean Beck Depression Inventory scores, adjusted for education and marital status, as in the log transformed analyses.
[a]Analysis for education was adjusted for marital status only, and analysis for marital satisfaction was adjusted for education only.
[b]Analysis conducted for the married women only.

TABLE 3—Mean (Log Transformed) Beck Depression Inventory Scores, Adjusted for Education and Marital Status, of Four Employment Groups

Examination	Always Employed	Always Nonemployed	Newly Nonemployed	Newly Employed
Baseline	.56 (4.2)	.61 (5.0)	.56 (4.1)	.71 (6.4)
3rd year	.60 (4.8)	.68 (6.4)	.60 (4.7)	.58 (5.9)

Note. The numbers in parentheses are the untransformed mean Beck Depression Inventory scores, adjusted for education and marital status.

these areas generated two factors; only one had good internal reliability ($\alpha = .76$) and included responsibilities for household chores, children, parents and relatives, and the happiness of the family.

Two types of social support were measured: perceived and structural. The former was determined from the appraisal support subscale of the Interpersonal Support Evaluation List, which has good internal and test-retest reliability.[23] This subscale assesses the subject's perception of the availability of someone to talk to about her problems. An alpha coefficient of .80 in our sample was found. Because of the skewed distribution of perceived support in this sample, an approximately one-third, two-thirds split was used for low and high support, respectively. Structural support consisted of two questions asking the subject how many (1) friends/neighbors and (2) family members not living with her she has kept in touch with during the previous 6 months? The variables were dichotomized in a one-quarter, three-quarters split for low and high support.

Using a modified version of the Pilkonis Life Event Schedule,[24,25] women indicated whether they had experienced any of 54 events in the 6 months prior to their 3-year clinic visit. They then indicated what the three most important events were and how stressful (not very, moderately, very) each one was for them. Stress was evaluated as a dichotomous variable: present (very stressful important event) or absent (no important events or not very/moderately stressful important event).

Results

Cross-Sectional Association between Employment Status and Depressive Symptoms

One-way analysis of covariance (ANCOVA) showed that, at study entry, nonemployed women had a significantly higher mean Beck Depression Inventory score than did employed women ($F[1, 520] = 5.44$; $P = .02$). A planned contrast procedure, in which a two (employment group: employed vs nonemployed) by two (low on factor vs high on factor) ANCOVA was followed by the calculation of a contrast F,[26] tested the hypothesis that women who reported a low level of education, perceived support, family support, friend support, or marital satisfaction (married only) and who were nonemployed would show higher Beck scores than all other women. All Fs were significant, thus confirming our hypothesis and replicating previous cross-sectional results (see Table 2).

Relationship between Change of Employment Status and Change in Depressive Symptoms

A planned contrast, in which a four (employment group: always employed, always nonemployed, newly nonemployed, and newly employed) by two (examination) repeated measures ANCOVA on the Beck scores measured at baseline and at 3-year follow-up was followed by the calculation of two contrast Fs, tested the hypothesis that the newly nonemployed would report an increase in symptoms from baseline to the 3-year exam, the newly employed would show a decrease, and both the always employed and the always nonemployed would report similar levels at both times. Results indicated an absence of support for our specific hypothesis.

Nonetheless, the initial ANCOVA showed a significant group-by-time interaction ($F[3, 520] = 3.31$; $P < .02$), indicating that the amount of change in depressive symptomatology over time differed among the four groups. The symptom levels increased over time for three of the groups while they decreased for one, the newly employed (Table 3). A follow-up repeated measures ANCOVA comparing the newly employed with the other three groups combined showed a significant group-by-time interaction ($F[1, 522] = 9.73$; $P < .007$ [Bonferroni adjusted]), suggesting that new employment was beneficial. This effect remained when the baseline Beck score was included as a covariate.

Influence of Psychosocial and Job Characteristics on Change in Depressive Symptoms

Planned contrasts also evaluated how the characteristics of the woman and

her job affected the relationship between change in employment status from baseline to 3-year exam and change in depressive symptom level. Because of the number of ANCOVAs conducted (n = 32), $P < .01$ was used as our significance level. Results were all nonsignificant.

Five characteristics of the women assessed at baseline did influence Beck scores, irrespective of employment status. Women's perception of low support, low family support, low marital satisfaction (married only), low job satisfaction, and negative job future at the beginning of the study were significantly associated with high symptom levels at the same time as well as 3 years later (Table 4). These findings were consistent with those found in previous literature.[10,27-29]

TABLE 4—Mean (Log Transformed) Beck Depression Inventory Scores, Adjusted for Education and Marital Status,[a] According to Baseline Psychosocial Characteristics

Characteristic	Examination	High	Low	*F* Value on High/Low Characteristic from Analysis of Covariance[c]
Perceived support	Baseline	.53 (3.8)	.76 (7.3)	
	3rd year	.53 (4.0)	.78 (8.2)	32.7
Family support	Baseline	.57 (4.2)	.74 (7.3)	
	3rd year	.57 (4.6)	.75 (8.0)	13.0
Marital satisfaction	Baseline	.43 (2.6)	.71 (6.3)	
	3rd year	.46 (3.3)	.72 (7.0)	38.7
Job satisfaction[b]	Baseline	.46 (2.9)	.67 (5.5)	
	3rd year	.53 (3.9)	.70 (5.9)	13.7
Job future[b]	Baseline	.44 (2.8)	.64 (5.1)	
	3rd year	.49 (3.4)	.70 (5.8)	16.4

Note. The numbers in parentheses are the untransformed Beck Depression Inventory scores, adjusted for education and marital status, as in log transformed analyses.
[a]Analysis for marital satisfaction was conducted for the married women and was adjusted for education only.
[b]Analyses conducted for those employed at baseline.
[c]All values are significant at $P < .001$.

TABLE 5—Mean (Log Transformed) Beck Depression Inventory Scores, Adjusted for Education and Marital Status, of Employment/Stress Groups

	Newly Employed		All Others	
Examination	No Stress	Stress	No Stress	Stress
Baseline	.67 (5.9)	.78 (7.4)	.52 (3.7)	.67 (5.5)
3rd year	.46 (4.7)	.84 (8.4)	.55 (4.1)	.76 (7.0)

Note. The numbers in parentheses are the untransformed mean Beck Depression Inventory scores, adjusted for education and marital status.

Life Stress and Employment

The well-established association between life stress and depressive symptoms and disorder[30-32] led us to examine if stress might explain the difference in Beck scores between the newly employed and the other women. The proportion of women who reported a stressful event in each employment group was approximately the same: 30%. Results of a two (newly employed vs all other women) by two (presence vs absence of stress) ANCOVA with repeated measures performed on the Beck scores showed a significant main effect for stress on symptoms ($F[1, 518] = 13.99$; $P < .01$). The three-way interaction between time, stress, and group was of borderline significance ($F[1, 520] = 2.90$; $P < .10$). Beck scores decreased for the newly employed women who did not report a stressful event but tended to increase for all the other groups (Table 5). Thus, our exploratory analyses suggest that the decrease in Beck scores for the newly employed women might be restricted to those not experiencing a stressful life event in the 6 months prior to their 3-year exam.

Discussion

Our cross-sectional analyses were consistent with those of other cross-sectional studies[5,33,34] in showing that women who work are psychologically better off than those who do not. Also consistent with previous cross-sectional research, nonemployed women with low perceived support, family support, friend support, education, or marital satisfaction were more symptomatic than those who were high on these characteristics. Unexpectedly, the longitudinal analyses did not reveal the same picture. Irrespective of the women's psychosocial characteristics at baseline, depressive symptoms decreased among the newly employed and increased for the others.

Exploratory analyses indicated that the newly employed women who reported no stress at the 3-year exam were the only subjects to show, on average, a decrease in depressive symptoms; all other women showed either no change or an increase in symptoms. Perhaps under conditions of low life stress, employment has the potential for enhancing mental health, whereas under stressful life circumstances, the positive effects of employment may be suppressed. Although a few studies have observed a significant relationship between work and/or family stresses and psychological symptomatology among employed women,[35,36] little is known about the combined and differential effects of various stressful events or situations for the employed and nonemployed women. Further studies are needed to examine specifically the interactive and additive effects of stress and employment/nonemployment on women's mental health.

The discrepancy between the cross-sectional results and those of our longitudinal analyses highlight the methodological limitations of cross-sectional research. An important advantage of our design is that the variables we examined as potential moderators of the employment-symptom relationship were assessed at least 2 years prior to any change in employment status. Thus, these variables could not be confounded with subsequent symptoms as they probably were with current ones. Although the contradictions between the cross-sectional and longitudinal results could be due to poor measurement or chance findings in the former, this is unlikely since our cross-sectional findings were highly significant and consistent with previous cross-sectional literature. Moreover, three of the measures used had previously established good reliability (per-

ceived support, educational level) or internal reliability (perceived support, marital satisfaction) in the study sample. Therefore, we conclude that the significant interactions between the independent variables and employment status obtained in the cross-sectional analyses probably reflect the characteristics of those women who are unable to seek and maintain employment and thus are not a true effect for employment.

Because our study was originally designed to assess the effects of menopause on cardiovascular risk factors, we had no information about why the women were or were not working, what their expectations of work were, and what they and their spouses thought about women who worked. We also had no information about the psychiatric histories of the women, an important determinant of current psychiatric distress and disorder. Nonetheless, we did have ample data—more than many studies—regarding other characteristics of the women and their work, and the data were prospective in nature.

In conclusion, new employment appears to have an acute positive effect on mood. Despite the many analyses we conducted and our own cross-sectional results, when examined prospectively, no subgroups seem to benefit more than others from new employment. Post hoc analyses suggest that for middle-aged, predominantly middle-class women, new employment might be especially beneficial among those who have not experienced a stressful life event in the previous 6 months. Consistent with the enhancement theory of multiple roles, employment is not detrimental to women's psychological well-being. □

Acknowledgments

This research was funded in part by National Institutes of Health grants HL 28266 and MH 15169.

We wish to thank Mary Amanda Dew, PhD, for her helpful statistical advice.

This paper is based in part on the dissertation project of the first author.

References

1. Matthews KA, Rodin J. Women's changing work roles: impact on health, family, and public policy. *Am Psychol.* 1989;44:1389–1393.
2. Foster CD, Siegel MA, Jacobs NR, eds. *Women's Changing Role.* Wylie, Tex: Information Aids, Inc; 1988.
3. Baruch GK, Barnett R. Role quality, multiple role involvement, and psychological well-being in midlife women. *J Pers Soc Psychol.* 1986;51:578–585.
4. Bebbington P, Hurry J, Tennant C, Sturt E, Wing JK. Epidemiology of mental disorders in Camberwell. *Psychol Med.* 1981;11:561–579.
5. Cochrane R, Stopes-Roe M. Women, marriage, employment and mental health. *Br J Psychiatry.* 1981;139:373–381.
6. Aneshensel CS, Frerichs RR, Clark VA. Family roles and sex differences in depression. *J Health Soc Behav.* 1981;22:379–393.
7. Pearlin LI. Sex roles and depression. In: Datan N, Ginsberg LH, eds. *Proceedings of Fourth Life-Span Developmental Psychology Conference: Normative Life Crises.* New York, NY: Academic Press; 1975:191–207.
8. Welch S, Booth A. Employment and health among married women with children. *Sex Roles.* 1977;3:385–397.
9. Warr P, Parry G. Depressed mood in working-class mothers with and without paid employment. *Soc Psychiatr.* 1982;17:161–165.
10. Kessler RC, McRae J Jr. The effect of wives' employment on the mental health of married men and women. *Am Sociol Rev.* 1982;47:216–227.
11. Krause N, Markides KS. Employment and psychological well-being in Mexican American women. *J Health Soc Behav.* 1985;26:15–26.
12. Vanfossen BE. Sex differences in the mental health effects of spouse support and equity. *J Health Soc Behav.* 1981;22:130–143.
13. Kandel DB, Davies M, Raveis VH. The stressfulness of daily social roles for women: marital, occupational and household roles. *J Health Soc Behav.* 1985;26:64–78.
14. Waldron I, Jacobs JA. Effects of multiple roles on women's health. Evidence from a longitudinal study. *Women Health.* 1989;15:3–19.
15. Ickovics MA, Martin J. The impact of employment on the psychological well-being of army wives: a longitudinal survey study. *Milit Med.* 1987;152:500–503.
16. Notman M. Midlife concerns of women: implications of the menopause. *Am J Psychiatry.* 1979;136:1270–1274.
17. Matthews KA, Kelsey SF, Meilahn EN, Kuller LH, Wing RR. Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. *Am J Epidemiol.* 1989;129:1132–1144.
18. Campbell RT, Hudson CM. Synthetic cohorts from panel surveys. *Res Aging.* 1985;7:81–93.
19. Beck AT, Beamesderfer A. Assessment of depression: the depression inventory. In: Pinchot P, ed. *Psychological Measurements in Psychopharmacology.* Basel, Switzerland: Karger Press; 1974:151–169.
20. Caplan RD, Cobb S, French JRP Jr, Van Harrison R, Pinneau SR Jr. *Job Demands and Worker Health: Main Effects and Occupational Differences.* Ann Arbor, Mich: Institute for Social Research; 1980.
21. Cottington EM. *Occupational Stress, Psychosocial Modifiers, and Blood Pressure in a Blue-Collar Population.* Pittsburgh, Pa: University of Pittsburgh, Graduate School of Public Health; 1983. Doctoral dissertation.
22. Haynes SG, Levine S, Scotch N, Feinleib M, Kannel WB. The relationship of psychosocial factors to coronary heart disease in the Framingham study: I. methods and risk factors. *Am J Epidemiol.* 1978;107:362–383.
23. Cohen S, Mermelstein R, Kamarck T, Hoberman HM. Measuring the functional components of social support. In: Sarason IG, Sarason B, eds. *Social Support: Theory, Research and Applications.* Dordrecht, the Netherlands: Martinus Nijhoff; 1985:73–94.
24. Pilkonis PA, Imber SD, Rubinsky P. Influence of life events on outcome in psychotherapy. *J Nerv Ment Dis.* 1984;172:468–474.
25. Pilkonis PA, Imber SD, Rubinsky P. Dimensions of life stress in psychiatric patients. *J Hum Stress.* 1985; Spring:5–10.
26. Rosenthal R, Rosnow RL. *Contrast Analysis: Focused Comparisons in the Analysis of Variance.* New York, NY: Cambridge University Press; 1985.
27. Turner RJ. Social support as a contingency in psychological well-being. *J Health Soc Behav.* 1981;22:357–367.
28. Waring EM, Patton D, Neron CA, Linker W. Types of marital intimacy and prevalence of emotional illness. *Can J Psychiatry.* 1986;31:720–725.
29. Williams AW, Ware JE Jr, Donald CA. A model of mental health, life events, and social supports applicable to general populations. *J Health Soc Behav.* 1981;22:324–336.
30. Brown GW, Bhrolchain MN, Harris T. Social class and psychiatric disturbance among women in an urban population. *Sociology.* 1975;9:225–254.
31. Brown GW, Bifulco A. Motherhood, employment and the development of depression: a replication of a finding? *Br J Psychiatry.* 1990;156:169–179.
32. Parry G. Paid employment, life events, social support, and mental health in working-class mothers. *J Health Soc Behav.* 1986;27:193–208.
33. Rosenfield S. Sex differences in depression: do women always have higher rates? *J Health Soc Behav.* 1980;21:33–42.
34. Aneshensel CS. Marital and employment role-strain, social support, and depression among adult women. In: Hobfall S, ed. *Stress, Social Support, and Women.* Washington, DC: Hemisphere Publishing Corp; 1986:99–114.
35. Frone MR, Russell M, Cooper ML. Relationship of work and family stressors to psychological distress: the independent moderating influence of social support, mastery, active coping, and self-focused attention. In: Perrewe PL, ed. *Handbook on Job Stress. J Soc Behav Pers.* 1991;6(special issue):227–250.
36. Phelan J, Schwartz JE, Bromet EJ, et al. Work stress, family stress and depression in professional and managerial employees. *Psychol Med.* 1991;21:999–1012.

Providing Medical Care to Methadone Clinic Patients: Referral vs On-Site Care

Annie Umbricht-Schneiter, MD, David H. Ginn, MD, Katherine M. Pabst, CRNP, MPH, and George E. Bigelow, PhD

ABSTRACT

Objectives. Intravenous drug users are at high risk for medical illness, yet many are medically underserved. Most methadone treatment programs have insufficient resources to provide medical care. The purpose of this study was to test the efficacy of providing medical care at a methadone clinic site vs referral to another site.

Methods. Patients with any of four target medical conditions were randomized into an on-site group offered medical care at the methadone treatment clinic and a referred group offered medical care at a nearby clinic. Entry to treatment and use of medical services were analyzed.

Results. Of 161 intravenous drug users evaluated, 75 (47%) had one or more of the target medical conditions. Fifty-one were randomized. In the on-site group (n = 25), 92% received medical treatment; in the referred group (n = 26), only 35% received treatment.

Conclusions. Providing medical care at a methadone treatment program site is more effective than the usual referral procedure and is a valuable public health intervention. (*Am J Public Health.* 1994;84:207–210)

Introduction

Intravenous drug users are at high risk for medical problems.[1–7] A substantial portion of this population consults physicians only for catastrophic health crises (mostly through emergency rooms); they also tend to show poor compliance with treatment.[8] Alternative approaches to health care delivery to intravenous drug users are needed. Integrating medical services into addiction treatment clinics has the potential for improving the prevention and treatment of diseases in this population,[9] and methadone treatment is effective in decreasing high-risk behaviors such as intravenous drug use and needle sharing. The acquired immunodeficiency syndrome (AIDS) epidemic has renewed public health interest in methadone treatment and in the need for delivering effective medical care to intravenous drug users. Some studies have shown that treatment of human immunodeficiency virus (HIV) infection or tuberculosis is feasible at methadone treatment clinics[10]; however, the relative efficacy of such interventions has not been documented.

Providing comprehensive medical care at a methadone treatment clinic offers an opportunity to screen for diseases and promote health education and behavioral changes. When patients come to the clinic to receive methadone, their compliance with other treatments can be promoted and/or directly supervised.

This study assessed the prevalence of four target medical conditions in a population of intravenous drug users seeking treatment for their addiction and compared the effectiveness of two methods for addressing these conditions: direct medical treatment at the addiction treatment clinic site (on-site care) and referral to a medical clinic (referred care), which is the usual procedure.

Methods

Recruitment, Eligibility, and Randomization

The study site was a hospital-based methadone treatment research clinic in southeastern Baltimore. The patient population consisted of opioid-dependent drug users; many were also using other drugs, mainly alcohol and cocaine. Patients received a medical examination on admission to methadone treatment. Medical care was free for the duration of the study if patients agreed to participate in a randomized trial that involved receiving medical care at the drug treatment site (on-site group) or being referred to a nearby medical clinic (referred group). Participants provided written informed consent. Those who needed further medical care for any of four target conditions and who had no primary care physicians were randomized. The target conditions were hypertension, purified protein derivative conversion (tuberculosis exposure), positive HIV serology (asymptomatic), and acute sexually transmitted diseases. At randomization, all patients were seen by the on-site medical staff and informed of their medical condition, need for treatment, and need for HIV posttest counsel-

The authors are with the Behavioral Pharmacology Research Unit, Johns Hopkins Medical School and Francis Scott Key Medical Center, Baltimore, Md.

Requests for reprints should be sent to George E. Bigelow, PhD, Behavioral Pharmacology Research Unit, Johns Hopkins/Key Medical Center, 5510 Nathan Shock Dr, Baltimore, MD 21224.

This paper was accepted June 21, 1993.

TABLE 1—Demographic, Social, and Diagnostic Characteristics of the Two Groups

	On-Site Group (n = 25), No.	Referred Group (n = 26), No.
Men	17	20
Non-Whites	16	19
Working status		
Working	5	6
Unemployed	17	17
Disabled	1	2
Marital status		
Single	11	10
Married	3	7
Divorced or separated	8	9
Insurance status		
Medical assistance	10	10
Private	2	7
None	13	8
Source of care		
Emergency room	19	17
Continuous care[a]	4	7
Reason for initial referral[b]		
HTN	4	3
PPD	7	9
HIV positive	7	4
STD	7	10
Distribution of diagnoses[b]		
HTN	4	3
PPD	8	14
HIV positive	7	4
STD	9	13

Note. HTN = hypertension, PPD = purified protein derivative, HIV = human immunodeficiency virus, STD = sexually transmitted disease.
[a]Continuous care by private physician, hospital clinic, or health maintenance organization.
[b]Patients could have several diagnoses but only one reason for initial referral.

ing. Patients randomized to the on-site group were treated immediately, or a follow-up appointment was arranged. The medical staff was actively involved in offering care to patients. Patients randomized to the referred group were instructed to contact a medical clinic located on the same campus as the addiction clinic, received written instructions describing the clinic location, and were given the telephone number of the clinic. The referred patients were responsible for obtaining an appointment at the medical clinic, where a physician was available twice a week; there was no waiting list, and appointments could be obtained within 4 days. To facilitate interactions with personnel at the medical clinic and ensure proof of coverage, all administrative information was collected by the research assistant at randomization and communicated to the clinic. To avoid economic confounds, the study covered the cost of medical care, including medications, for 3 months after randomization.

Patients were interviewed by a research assistant at 4 and 8 weeks after randomization. If a patient had not received medical treatment by week 4, the on-site medical staff stressed the importance of such treatment. A patient remaining untreated at week 8 was considered a treatment failure. For those patients still enrolled in drug treatment, medical care was then offered on-site. This "rescue treatment," of which patients were unaware, was arranged for ethical reasons. (Rescue treatment was offered to patients with acute sexually transmitted diseases at week 4).

At the end of the study, patients were referred to medical clinics elsewhere. Patients without insurance at enrollment were assisted in applying for medical assistance or other benefits.

Medical Evaluation

Hypertension. Patients with a diastolic blood pressure of 100 mm Hg or more were randomized. Patients with a diastolic blood pressure of 95 to 100 mm Hg were randomized if their blood pressure remained elevated after the first week of drug treatment.

Tuberculosis exposure. Patients with no prior history of tuberculosis exposure were tested with five standard tuberculin units (by intradermal injection) and examined 48 hours later. A reaction was considered positive if an induration of 5 mm or more was noted. (This represents the Centers for Disease Control recommendation for individuals who are HIV positive; because HIV status was unknown, these guidelines were extended to all patients.)[11,12]

Sexually transmitted diseases. Rapid plasma reagin (RPR) was obtained. In men, a first-stream urine was tested for leukocyte esterase and pyuria (more than five white blood cells per high-power field). Patients with positive urine were randomized, and a urethral sample was then analyzed for *Neisseria gonorrhoeae* and *Chlamydia trachomatis* by DNA probe.[13] Women were screened with a Pap smear and tested for *Candida albicans, Trichomonas vaginalis, N. gonorrhoeae,* and *C. trachomatis.* Although we strongly stressed the importance of the pelvic examination, women were allowed to refuse it.

HIV. Written informed consent was obtained after pretest counseling. If patients accepted testing, their blood sample was tested confidentially. Patients with symptomatic HIV infection (AIDS-related complex, AIDS) were referred to specialized clinics for care and were not randomized.

Addiction Treatment

Patients were enrolled in combined pharmacological and counseling treatment protocols for their drug addiction. The details of these protocols are beyond the scope of this paper.

Data Collection

Medical charts of newly admitted patients were reviewed weekly. A flow sheet was maintained to record the medical conditions and treatment assignment. The patient's counselor received a portion of the flow sheet to help coordinate addiction and medical treatment.

Enrollment in medical care was recorded, by patient interview and chart review, at 4 and 8 weeks after randomization. Primary outcome measures were number of visits per patient (visits for other medical conditions were not counted) and whether patients complied with the treatment plan (treatment rate per diagnosis).

Statistical Analysis

Comparisons between the two groups were made with *z* tests for proportions on dichotomous variables; *t* tests were used on continuous variables.

Results

Prevalence of the Target Medical Conditions

Of 161 patients evaluated, 75 had one or more of the target conditions: 12 had hypertension, 26 evidenced purified protein derivative conversion, 24 had positive HIV serology (27 [18%] refused testing), and 29 had sexually transmitted diseases (14 [25%] of 57 women refused pelvic examination).

Patient Characteristics

The characteristics of the 51 patients who were randomized are shown in Table 1. There were no significant differences between the two groups on any of the listed variables or in addiction treatment modality.

Enrollment in Medical Treatment

Of 25 patients randomized on-site, 23 (92%) enrolled in medical treatment and

complied with medical recommendations. Of 26 patients referred off-site, only 9 (32%) enrolled in medical treatment (8 enrolled at the assigned off-site clinic and 1 called a physician for a prescription). The difference between the two groups was statistically significant ($P < .001$).

Table 2 describes the medical care received by the two groups. All patients in the on-site group were seen at least once regarding their medical condition(s), 23 accepted treatment, and 19 (76%) had one or more additional follow-ups. The average number of visits per patient was 3.1 (SD = 1.8); the treatment rate per diagnosis was 88%. In the referred group, 8 patients (31%) were seen once at the off-site medical clinic and 1 obtained treatment elsewhere; of 5 patients scheduled for follow-up, 1 returned. The treatment rate per diagnosis was 28%.

The greater overall efficacy of the on-site procedure was corroborated in analyses within each of the four individual target medical conditions.

Hypertension. Four patients in the on-site group were successfully treated with normalization of their blood pressure (three with behavioral treatment, one with pharmacological treatment). Of the three patients in the referred group, two were started on antihypertensive therapy but did not return.

Tuberculosis exposure. Of 8 patients in the on-site group, 1 remembered prior tuberculosis treatment as a child. Of the 7 remaining patients, 6 obtained a chest x-ray. Chemoprophylaxis was indicated and administered at the methadone window to 1 of them. In spite of three medical visits, 1 patient did not comply with the request for chest x-ray. Of 14 patients in the referred group, 1 had a history of previous treatment. Of the remaining 13, only 3 obtained a chest x-ray. The rescue treatment allowed an additional 3 patients to be evaluated on-site, and 1 of them received chemoprophylaxis.

HIV. Of seven patients in the on-site group, five gave blood for the CD4 cell count, and two patients meeting criteria for retroviral therapy received zidovudine (AZT). Two patients refused further evaluation; one of them was seen seven times for intensive counseling, and both were referred to HIV clinics. Of the four patients in the referred group, only one came for the CD4 cell count. This patient's follow-up was interrupted involuntarily (by incarceration).

Sexually transmitted diseases. Of 9 patients randomized to the on-site group, 8 were treated successfully (1 twice, after reexposure). One patient with secondary syphilis completed appropriate therapy, and 1 patient with treated syphilis did not need further care. Of the 13 patients randomized to the referred group, 4 received treatment. Another patient with treated syphilis did not require treatment. Two other patients received rescue treatment on-site.

TABLE 2—Treatment Information for the Two Groups

	On-Site Group (n = 25)	Referred Group (n = 26)	P
No. of patients seen one or more times	25	8	<.001
No. of patients seen two or more times	19	1	<.001
No. of visits per patient, mean ± SD	3.1 ± 1.8	0.4 ± 0.6	<.001
Treatment rate, %[a]	88	28	<.001

[a]The denominator is the number of diagnoses rather than the number of patients.

Discussion

More than 45% of intravenous drug users seeking methadone treatment presented with a condition requiring medical care. When on-site medical care was available, more than 90% received treatment; when referred medical care was available, only 35% received treatment.

Prevalence of Medical Conditions

Among intravenous drug users seeking addiction treatment, we found that half had a condition, often unknown to them, that required medical care. This is a conservative assessment, because our study recorded only four target medical conditions. Three of these conditions (tuberculosis, sexually transmitted diseases, and HIV infection) currently represent serious public health threats because they reach epidemic proportions not only among intravenous drug users but also among their close contacts. Because of their asymptomatic course, it is important that they be detected and treated in a timely and effective manner.

Although not reported in this study, we found that many more patients suffered from other medical conditions such as soft-tissue infections, liver damage, respiratory diseases, and anemia. For ethical reasons, these conditions were treated by the on-site medical staff.

Efficacy of On-Site Care

In the on-site group, 92% of intravenous drug users entering addiction treatment received medical care, with a treatment rate per diagnosis of 88%. In comparison, in the referred group (with optimal referral conditions), only 35% of patients received medical care for the same conditions, with a treatment rate per diagnosis of 28%.

The difference in enrollment for medical care between the two groups may reflect, in part, a difference in providers' behavior rather than patients' behavior. In the drug treatment program, the medical team used nontraditional doctor–patient approaches: accepting unscheduled visits (initiated when "spotting the provider in the hall") and rescheduling several times for missed appointments. Coordination with the nurses dispensing methadone was of prime importance: they reminded patients of their medical appointment and, when needed, "held" the methadone temporarily until patients reported to the medical provider. Furthermore, compliance with medical treatment could be fostered by administering other medications along with the methadone. No patient was coerced into accepting medical treatment, but avoiding interaction with the medical staff was difficult.

We noted that patients used the on-site medical services extensively for other conditions or for health education. This suggests that intravenous drug users are interested in health care when it is available.

The presence of medical providers at the addiction treatment clinic site alleviated some of the burden usually carried by nurses or counselors alone. These providers assisted in the evaluation of acutely ill patients and reduced the need for referral to local emergency rooms.

Providing medical care to intravenous drug users remained a challenge, even at an addiction treatment clinic site; however, the study showed that this type of health care delivery resulted in high enrollment in medical care and an excellent treatment rate, whereas the usual referral process resulted in an unacceptable loss of patients to medical care.

While the emphasis of the present discussion is that methadone clinics can

enhance medical treatment efficacy, it should be noted that recent research[14] also suggests the converse: providing medical and psychosocial services can enhance the efficacy of addiction treatment. Given the current infectious disease epidemics, intravenous drug users should have increased access to methadone treatment providing comprehensive care.

This study is, to our knowledge, the first controlled randomized trial to test the efficacy of providing primary medical care at an addiction treatment site. The high prevalence of medical conditions in this population may be sufficient rationale, at least on a humanitarian basis, to provide medical care at addiction treatment clinics. The superior efficacy of this approach, as documented here, further strengthens this rationale. Additional study is needed (e.g., assessments of emergency room visits or hospitalizations) to evaluate the cost-effectiveness of providing medical care at addiction treatment sites.

Barriers to Referred Medical Care

For several reasons, intravenous drug users are unlikely to use traditional medical services. Many fear discriminatory or hostile behavior from the medical staff in general clinics or hospitals.[15,16] If the drug users are hospitalized, their methadone treatment may be discontinued and the resulting withdrawal symptoms not adequately treated.

Many fear the stigma of being labeled HIV infected if seen in an addiction treatment clinic.

Medical practices tend to avoid referrals from drug treatment clinics. Four medical practices were asked to serve as referral sites for this study, and, although occasional individual physicians (with some training in addiction medicine) agreed to participate, all four practices refused. This was especially notable given that the study guaranteed fee payment, including a fee for "missed appointments." This reluctance to provide care to drug users appeared related to concerns about the behavioral stereotypes of the population, as well as possible economic risks after the conclusion of the study.

For intravenous drug users entering drug abuse treatment, the active planning required to obtain health care (such as scheduling an appointment and arranging for timely transportation) involves a level of structured behavior that is a goal of, rather than a prerequisite for, addiction treatment. Medical and addiction clinics rarely coordinate opening hours, thus complicating the patient's task further.

If drug treatment clinics are to provide screening for diseases such as tuberculosis, HIV infection, and sexually transmitted diseases (as federal guidelines currently recommend), they should also be able to provide medical treatment; issues of confidentiality restrict communication between providers, and patients often misunderstand their medical conditions.

Finally, this study featured an optimal referral process; detailed oral and written information on how to contact the clinic was provided, and all monetary obstacles were removed. Still, only 35% of the referred group enrolled in medical care. In reality, compliance with the usual referral procedure is probably less successful, because patients may lack adequate health care coverage (in our program, 40% of the patients had no insurance).

Conclusions

This study demonstrated a high prevalence of concurrent medical conditions requiring further care among intravenous drug users entering addiction treatment. When medical care was available on-site at the addiction treatment clinic, nearly all received care. When medical care was available through the usual referral procedure, few received care. These data argue for the effectiveness and public health value of providing medical care resources within addiction treatment clinics. □

References

1. Cooper JR. Methadone treatment and acquired immunodeficiency syndrome. *JAMA.* 1989;262:1664–1668.
2. Haverkos HW, Lange WR. Serious infections other than human immunodeficiency virus among intravenous drug abusers. *J Infect Dis.* 1990;161:894–902.
3. Lamb RJ, Yeh IT. Tuberculosis, AIDS, and IV drug abuse. *NJ Med.* 1990;87:413–415.
4. Drucker E. AIDS and addiction in New York City. *Am J Drug Alcohol Abuse.* 1986;12:165–181.
5. Haverkos HW. HIV infection among intravenous drug abusers in the United States and Europe. *NIDA Res. Monogr.* 1988;80: 7–17.
6. Hahn RA, Onorato IM, Jones TS, Dougherty J. Prevalence of HIV infection among intravenous drug users in the United States. *JAMA.* 1989;261:2677–2684.
7. Uchtenhagen A. AIDS Epidemiologie und Pravention bei i.-v.-Drogenabhangigen. *Soz. Praventivmed.* 1988;33:326–330.
8. Dans PE, Matricciani RM, Otter SE, Reuland DS. Intravenous drug abuse and one academic health center. *JAMA.* 1990;263: 3173–3176.
9. Haverkos HW. Infectious diseases and drug abuse: prevention and treatment in the drug abuse treatment system. Presented at the European Symposium on AIDS and Drug Abuse, August 1990, Vienna, Austria.
10. Selwyn PA, Feingold AR, Iezza A, et al. Primary care for patients with human immunodeficiency virus (HIV) infection in a methadone maintenance treatment program. *Ann Intern Med.* 1989;111:761–763.
11. Advisory Committee for the Elimination of Tuberculosis. Tuberculosis and human immunodeficiency virus infection. *MMWR.* 1989;34:236–250.
12. Centers for Disease Control. Screening for tuberculosis and tuberculous infection in high-risk populations. *MMWR.* 1990;39: 1–7.
13. Mercer LJ, Robinson DC, Sahm DF, Lawrie MJ, Hajj SN. Comparison of chemiluminescent DNA probe to cell culture for the screening of chlamydia trachomatis in a gynecologic clinic population. *Obstet Gynecol.* 1990;76:114–117.
14. McLellan AT, Arndt IO, Metzger DS, Woody GE, O'Brien CP. The effects of psychosocial services in substance abuse treatment. *JAMA.* 1993;269:1953–1959.
15. Schindler BA, Blum D, Malone R. Noncompliance in the treatment of endocarditis. The medical staff as co-conspirators. *Gen Hosp Psychiatry.* 1988;10:197–201.
16. McGrory BJ, McDowell DM, Muskin PR. Medical students' attitudes toward AIDS, homosexuals and intravenous drug abusing patients: a re-evaluation in New York City. *Psychosomatics.* 1990;31:426–433.

The Relationship between Mortality and Intensity of Inpatient Alcoholism Treatment

ABSTRACT

Objectives. Previous studies have examined mortality in alcoholics receiving extended inpatient alcoholism treatment, but few have investigated less intense treatment. This study examined mortality within 3 years after discharge from varying intensities of inpatient alcoholism treatment.

Methods. Using the computerized database of the Department of Veterans Affairs, we identified men participating in varying intensities of inpatient alcoholism treatment and followed them for 3 years after discharge. Adjusted mortality rates were computed and survival analysis was performed to assess the risk of death, adjusting for factors that may be related to mortality.

Results. The death rate was lower for men who completed extended formal inpatient treatment than for those who began, but did not complete, inpatient treatment or those who underwent short detoxification. Differences among the treatment groups remained after age, race, marital status, and disease severity were controlled.

Conclusions. These results suggest that extended formal inpatient alcoholism treatment is associated with a lower risk of death than less intense forms of inpatient treatment. (*Am J Public Health.* 1994;84:211–214)

Janice Y. Bunn, PhD, Brenda M. Booth, PhD, Cynthia A. Loveland Cook, PhD, Frederic C. Blow, PhD, and John C. Fortney, PhD

Introduction

The increased risk of premature death among alcoholics has been well documented;[1,2] potential life lost has been estimated to be about 15 years.[3] The risk of mortality for alcoholics has been reported to be two to three times that of the general population,[4,5] with more than sevenfold increases in mortality in some groups.[6,7]

Several studies have used hospital admission data to identify potential alcoholics. These patients, however, were not exclusively limited to those being treated for alcoholism. Rather, such analyses have included patients discharged from mental hospitals[5,8] or psychiatric units of general hospitals[6] with diagnoses indicating alcoholism but not necessarily representing formal alcoholism treatment. Other studies have included persons whose charts indicated the presence of alcohol-related problems but who were hospitalized for medical or surgical reasons.[9] As a result, the subsequent mortality of these patients may have been influenced by the medical diagnosis for which the patient was admitted, the number and severity of medical problems, the type of care provided during the hospitalization, or the type of hospital from which the patient was enrolled.[9,10]

Although alcoholism treatment that is less intense than extended inpatient treatment is available, patients receiving such care are often not included in follow-up studies. Thus, relatively little is known about mortality in treated alcoholics participating in programs of varying lengths and intensities of treatment. In this study we examined the mortality of a large sample of alcoholics who completed extended formal inpatient alcoholism treatment or less intense forms of inpatient alcoholism treatment. A large group of alcoholic patients hospitalized for medical, surgical, or psychiatric problems was included for comparison.

Methods

Identification of the Sample

The subjects in this study were identified through the Department of Veterans Affairs (VA) computerized database, the Patient Treatment Files. Each file contains discharge information for all patients treated at any VA facility in the country during a given fiscal year.

Each alcoholic was placed into one of three treatment groups on the basis of treatment received during fiscal year 1987: (1) patients who completed extended formal inpatient alcoholism treatment (n = 28 060); (2) patients who underwent at least 6 days of extended formal inpatient alcoholism treatment but who did not complete the treatment (n = 10 922); or (3) patients admitted for short detoxification lasting 5 days or less (n = 9424). A fourth group of hospitalized alcoholics who received no direct alcoholism treat-

At the time of the study, Janice Y. Bunn, Brenda M. Booth, and John C. Fortney were with the Department of Veterans Affairs Medical Center, Iowa City, Iowa. Drs Bunn and Booth were also with the Center for Health Services Research, The University of Iowa, Iowa City. Cynthia A. Loveland Cook was with the George Warren Brown School of Social Work, Washington University, St Louis, Mo. Frederic C. Blow is with the Department of Psychiatry, University of Michigan Alcohol Research Center, Ann Arbor, Mich.

Requests for reprints should be sent to Brenda M. Booth, PhD, Department of Psychiatry, University of Arkansas for Medical Sciences, 4301 W Markham, Mail Slot 554, Little Rock, AR 72205.

This paper was accepted July 26, 1993.

TABLE 1—Number of Deaths and Crude and Adjusted Mortality Rates for the 3-Year Period after Index Hospitalization for Alcoholism Treatment

	No. Patients	No. Deaths	Crude Mortality Rate[a]	Adjusted Mortality Rate[b]
Complete treatment	28 060	1982	24.35	23.23
Incomplete treatment	10 922	1030	32.96	29.30
Short detoxification	9 434	1057	39.60	35.12

[a]Deaths per 1000 person-years.
[b]Age- and race-adjusted deaths per 1000 person-years.

TABLE 2—Risk Ratios and 95% Confidence Intervals for Survival in the 3 Years after Index Hospitalization for Alcoholism Treatment

	Risk Ratio	95% Confidence Interval
Treatment group		
Complete treatment	1.00	
Incomplete treatment	1.24	1.14, 1.33
Short detoxification	1.44	1.33, 1.55
Age		
<30 y	1.00	
30–39 y	1.97	1.55, 2.50
40–49 y	3.20	2.52, 4.06
50–59 y	5.89	4.67, 7.44
60–69 y	8.96	7.10, 11.36
70+ y	10.12	7.66, 13.37
Disease severity		
Low	1.00	
Medium	1.22	1.13, 1.32
High	1.89	1.75, 2.04
Marital status		
Not married	1.00	
Married	0.75	0.70, 0.81
Race		
Caucasian	1.00	
Non-Caucasian	0.98	0.92, 1.05

ment was identified for comparison purposes (n = 26 261). These patients had primary diagnoses that were medical or psychiatric in nature and secondary diagnoses indicating chronic alcoholism. Details regarding sample selection are published elsewhere.[11] Because of the small number of female alcoholics treated in the VA system, the study population was limited to men.

Information on patients not using VA services at the end of 3 years, including name, social security number, date of birth, gender, and state of last residence, was sent to the National Death Index for comparison with the Index's files. The National Death Index maintains a computerized database of deaths recorded by state vital statistics offices and submitted to the National Center for Health Statistics. Data received from the Index were screened to identify matches for patients who died within 3 years of discharge from the index hospitalization. Any patients not identified through the Index records were assumed to be alive at the end of the 3-year follow-up period.

Data Analysis

The number of deaths and the number of years each person was alive following the treatment episode were tabulated to compute crude and adjusted mortality rates. Age-specific mortality rates were calculated for each racial group by dividing the number of deaths in a given age group by the total person-years in that age group. Mortality rates were then adjusted directly,[12] with the entire male population of the United States enumerated in the 1990 census used as the standard population.[13] Age-adjusted, race-specific rates were initially computed; however, because the age-adjusted mortality rates for Caucasians and non-Caucasians did not differ, they were combined to provide age- and race-adjusted mortality rates.

A survival analysis examined the time to death for the three treatment groups, controlling for potential confounding variables, including age, marital status, race, and severity of alcohol-related complications. Age was measured in 10-year intervals; marital status was coded as married or not married. Because of the relatively small number of individuals in some of the racial/ethnic groups examined, all non-Caucasians were combined. The severity of medical-related complications of alcoholism was determined by an approximation of the disease staging methodology,[14] allowing categorization into one of three stages of disease, ranging from disease with minimal severity and no complications to disease involving multiple physiological systems. The reference group was composed of unmarried Caucasian men younger than 30 years of age who completed extended formal inpatient treatment.

Results

More than 8% of the treated alcoholics (4069/48 416) died within 3 years following discharge from the index treatment episode. As described in Table 1, 7.1% of the group that completed extended formal alcoholism treatment died. Mortality in the incomplete treatment group (9.5%) and the short detoxification group (11.2%) was higher than for those who completed extended formal treatment. In the group of alcoholic patients who received no alcoholism-specific treatment in fiscal year 1987, 23% died.

The crude death rates ranged from 24.35 per 1000 person-years for the completed treatment group to 39.60 per 1000 person-years for those who underwent short detoxification (Table 1). Similarly, the age- and race-adjusted mortality rate increased with a decreasing amount of alcoholism-specific treatment and was highest in the group that received the least treatment, the short detoxification group (35.12 deaths per 1000 person-years). In contrast, those who received no direct alcoholism treatment had an age- and race-adjusted mortality rate of 56.72 deaths per 1000 person-years.

Results of the survival analysis for patients in the three treatment groups are presented in Table 2. The higher mortality rates observed in the less intense alcoholism treatment groups presented in Table 1 were reflected as increased risk ratios when these groups were compared with those who completed extended formal inpatient treatment. Men in the incomplete treatment group had a 24% elevation in mortality risk, whereas the risk ratio in the short detoxification group was 1.44.

There was an increasing risk of death with increasing age. Men 70 years of age and older experienced a greater than 10-fold increase in the risk of dying during the 3-year period. The presence of complications of alcoholism, as measured by disease severity, increased the risk of dying by 22% for those whose complications were defined as moderate and by almost 90% for those with more severe complications. Men who were married were less likely to die than those who were not (risk ratio of 0.75). The risk ratio for non-Caucasians in this study was not elevated above that for Caucasians.

The survival analysis was repeated to include the group of patients who received

no alcoholism-specific treatment. The risk ratios for all variables in the original model were essentially unchanged (data not shown). This group of individuals, however, exhibited a risk ratio of 2.19, relative to those who completed extended formal inpatient treatment, after the other factors in the model were controlled.

Discussion

This study is the largest known effort to examine mortality in alcoholics and one of the first to compare mortality in patients involved in a variety of inpatient treatment options. The large sample size used in this study made it possible to compute age- and race-adjusted mortality rates in the groups studied. These results verify the well-known increase in mortality in alcoholics compared with the general population. The adjusted mortality rate for men completing extended formal inpatient alcoholism treatment was twice that of the general population in 1989.[15] These results are similar to those reported by Schmidt and de Lint[4] and Barr et al.[16] for other groups of treated alcoholics. Neither of these studies, however, was able to capture the increased mortality associated with not completing the full course of inpatient treatment.

Previous work has shown that the risk of death following alcoholism treatment is significantly lower in those individuals who remain abstinent.[16,17] Although we cannot ascertain from our databases which or how many of the men undergoing treatment were successfully able to control their drinking, completing extended formal inpatient treatment was associated with significantly lower mortality compared with persons who received less intense alcoholism-specific treatment. Other researchers have recently found extended inpatient treatment to be associated with better patient outcomes than self-help groups.[18]

Although one would like to ascribe the decreased mortality among those who complete extended formal inpatient treatment to the alcoholism treatment itself, the issue of patients' "self-selecting" to treatment cannot be ignored. However, the incomplete treatment group did provide some control over self-selection.

Other studies of mortality in alcoholics identified through hospital records have grouped those undergoing alcoholism treatment with alcoholics who were hospitalized for other psychiatric and/or medical reasons.[5,6,8,9] This study substantiates the suggestion that these groups are clearly different.[10] Although we have previously shown that the men who received no alcoholism-specific treatment had more severe alcohol-related medical complications,[19] these differences in group characteristics did not completely explain the group's excess mortality. On the other hand, these factors may not be sufficient to explain all of the differences between those who were hospitalized but did not receive directed alcoholism treatment and those who were being treated specifically for alcoholism.

It is possible that patients in the less intense treatment groups received formal alcoholism treatment in years other than the index treatment year. They may also have received treatment from non-VA facilities. Alcoholics who receive VA treatment, however, are largely unemployed, with limited financial resources,[20] and are unlikely to be able to afford inpatient treatment outside the VA. The inclusion of persons who completed additional treatment would have the effect of attenuating the differences between those who completed treatment and the other groups. The reduction in mortality seen in those who completed alcoholism treatment, therefore, would be even more striking.

Taylor and colleagues[6] have suggested that the mortality of female alcoholics is similar to that of their male counterparts. Given the nature of the veteran population, there were too few women alcoholics to include in this study. Thus, we were unable to examine this issue.

Although the prevalence of alcoholism is similar in Caucasian and African-American groups,[21,22] the age-adjusted rate for alcohol-related deaths in non-Caucasian males was much higher than that of Caucasian males for the US population in 1989.[15] Herd also reported higher mortality from cirrhosis among Blacks, particularly in the younger age groups.[23] On the other hand, Mendenhall and colleagues[24] found that survival after diagnosis with alcoholic liver disease in male patients treated at VA medical centers was better for African Americans than for Caucasians, and that other non-Caucasian men fared worse than either of these groups. The higher overall mortality in non-Caucasian males has been attributed to differing racial distributions of socioeconomic status.[25] Because of income eligibility requirements for the use of VA services, these veterans have a relatively uniform lower socioeconomic status. Thus, race in this study may have contributed little excess risk beyond that already accounted for by low socioeconomic status.

It is likely that some alcoholics may have already begun experiencing some of the medical and/or psychiatric complications of alcoholism when they began treatment. Alcoholics with more severe complications may also be high users of both inpatient and outpatient services, suggesting a need to examine the possible relationship between mortality and health care utilization.

The provision of alcoholism treatment has been shown to be beneficial in reducing health care costs.[26] The results of this study, notwithstanding the limitations discussed above, suggest that more intense treatment modalities are also associated with more favorable mortality outcomes. □

Acknowledgments

This work was supported by grants RO1-AA08732 (all authors) and P50-AA07378 (Dr Blow) from the National Institute on Alcohol Abuse and Alcoholism.

References

1. Taylor JR, Helzer JE. The natural history of alcoholism. In: Kissin B, Begleiter H, eds. *The Pathogenesis of Alcoholism: Psychosocial Factors*. New York, NY: Plenum Press; 1983.
2. Vaillant GE. *The Natural History of Alcoholism*. Cambridge, Mass: Harvard University Press; 1983.
3. Haberman PW, Natarajan G. Premature mortality and chronic alcoholism: medical examiner cases, New Jersey. *Soc Sci Med*. 1989;29:729–732.
4. Schmidt W, de Lint J. Causes of death in alcoholics. *Q J Stud Alcohol*. 1972;33:171–185.
5. Nicholls P, Edwards G, Kyle E. Alcoholics admitted to four hospitals in England, II: general and cause-specific mortality. *Q J Stud Alcohol*. 1974;35:841–855.
6. Schuckit MA, Gunderson EKE. Deaths among young alcoholics in the U.S. Naval Service. *Q J Stud Alcohol*. 1974;35:856–862.
7. Finney JW, Moos RH. The long-term course of treated alcoholism, I: mortality, relapse and remission rates and comparisons with community controls. *J Stud Alcohol*. 1991;52:44–54.
8. Gorwitz K, Bahn A, Warthen FJ, Cooper M. Some epidemiological data on alcoholism in Maryland: based on admissions to psychiatric facilities. *Q J Stud Alcohol*. 1970;31:423–443.
9. Taylor JR, Combs-Orme T, Taylor DA. Alcohol and mortality. Diagnostic considerations. *J Stud Alcohol*. 1983;44:17–25.
10. Combs-Orme T, Taylor JR, Robins LN, Holmes SJ. Differential mortality among alcoholics by sample site. *Am J Public Health*. 1983;73:900–903.
11. Booth BM, Blow FC, Cook CAL, Bunn JY, Fortney JC. Age and ethnicity among

hospitalized alcoholics: a nationwide study. *Alcohol Clin Exp Res.* 1992;16: 1029–1034.
12. Lilienfeld AM, Lilienfeld DE. *Foundations of Epidemiology.* New York, NY: Oxford University Press; 1980:76–78.
13. *Census of Population and Housing, 1990. Summary Tape File 1 on CD-ROM.* Washington, DC: US Bureau of the Census; 1991.
14. Gonella JD, Hornbrook MC, Louis DZ. Staging of disease: a case-mix measure. *JAMA.* 1984;251:637–644.
15. National Center for Health Statistics. Advance Report of Final Mortality Statistics, 1989. *Month Vital Stat Rep.* 1992; 40(8)(suppl 2).
16. Barr HL, Antes D, Ottenberg DJ, Rosen A. Mortality of treated alcoholics and drug addicts: the benefits of abstinence. *J Stud Alcohol.* 1984;45:440–452.
17. Bullock KD, Reed RJ, Grant I. Reduced mortality risk in alcoholics who achieve long-term abstinence. *JAMA.* 1992;267: 668–672.
18. Walsh DC, Hingson RW, Merrigan DM, et al. A randomized trial of treatment options for alcohol-abusing workers. *N Engl J Med.* 1991;325:775–782.
19. Cook CAL, Booth BM, Blow FC, Gognem A, Bunn JY. Alcoholism treatment, severity of alcohol-related medical complications, and health services utilization. *J Ment Health Administration.* 1992;19:31–40.
20. McLellan AT, Luborsky L, O'Brien CP, Woody GE, Druley KA. Is treatment for substance abuse effective? *JAMA.* 1982; 247:1423–1428.
21. Robins LN, Helzer JE, Weissman MM, et al. Lifetime prevalence of specific psychiatric disorders in three sites. *Arch Gen Psychiatry.* 1984;41:949–958.
22. Herd D. Subgroup differences in drinking patterns among black and white men: results from a national survey. *J Stud Alcohol.* 1990;51:221–232.
23. Herd D. Migration, cultural transformation and the rise of black liver cirrhosis mortality. *Br J Addict.* 1985;80:397–410.
24. Mendenhall CL, Gartside PS, Roselle GA, Grossman CJ, Weesner RE, Chedid A. Longevity among ethnic groups in alcoholic liver disease. *Alcohol Alcohol.* 1989; 24:11–19.
25. Keil JE, Sutherland SE, Knapp RG, Tyroler HA. Does equal socioeconomic status in Black and White men mean equal risk of mortality? *Am J Public Health.* 1992;82: 1133–1136.
26. Holder HD, Blose JD. The reduction of health care costs associated with alcoholism treatment: a 14-year longitudinal study. *J Stud Alcohol.* 1992;53:293–302.

Completion Rates of Clients Discharged from Drug and Alcohol Treatment Programs in Washington State

ABSTRACT

Objectives. The primary goal of this study was to analyze completion rates of clients in drug and alcohol abuse treatment programs in Washington State and to assess the factors associated with treatment completion. A secondary goal was to examine the utility of a state information system as a source of evaluative data.

Methods. Analyses were conducted of 5827 client records contained in the Washington State Substance Abuse Monitoring System, representing a census of public clients discharged during the last quarter of 1990 from all state-funded alcohol and drug treatment programs in four treatment modalities. Logistic regression was performed to determine the independent predictors of treatment completion.

Results. Completion rates were highest for intensive inpatient alcohol treatment (75%) and lowest for intensive outpatient drug programs (18%). Factors associated with treatment completion included screening at a referral assessment center, education, age, ethnicity, and existence of a secondary drug problem.

Conclusions. The fit between clients and treatment programs may be an important factor explaining why some clients complete treatment and others drop out. State client information systems are an important source of data for analyzing treatment completion and other outcomes. (*Am J Public Health.* 1994;84: 215–221)

Thomas Wickizer, PhD, Charles Maynard, PhD, Adam Atherly, MA, Margaret Frederick, MPH, Thomas Koepsell, MD, Antoinette Krupski, PhD, and Kenneth Stark, MBA, MED

Introduction

Drug abuse and alcoholism represent significant public health problems that have generated increased concern in recent years. In 1989, an estimated $4 billion was spent to support drug and alcohol treatment programs in the United States,[1] almost half of which represented public expenditures. A sizable portion of these expenditures comes from states, which, in the face of federal cutbacks, have had to shoulder a greater burden to support needed treatment programs. Along with this burden has come a greater need for accountability and program monitoring.

These factors have prompted renewed interest in information systems that gather client and program data. Significant resources, at both the federal and state levels, have been devoted to the development of client-based data systems with mixed results. Many systems have had ongoing problems, including reporting delays, incomplete and inaccurate data input, and limited report-generating capabilities. To improve program and client monitoring for drug and alcohol treatment in Washington State, the State Division of Alcohol and Substance Abuse developed the Substance Abuse Monitoring System, a client-based statewide management information system.

In early 1991, the Division of Alcohol and Substance Abuse initiated a project to assess completion rates of state-funded drug and alcohol treatment programs in Washington based on analysis of data from the monitoring system. The project was conducted by researchers at the University of Washington School of Public Health during a 6-month period beginning in April 1991. The purpose of this paper is twofold: (1) to report the results of the analysis and (2) to demonstrate the potential of a state information system to provide data useful for evaluation and analysis. This study is notable because it is based on data representing a large and diverse client population that includes both youth and adult clients, as well as groups for whom treatment completion is rarely analyzed and reported in the literature, including Native Americans, Hispanics, and Asians.

Literature Review

Methodological Problems

Past studies have produced disparate results, probably as a result of differences in subject populations, methods and measures, and definitions of treatment completion. Definitions of completion vary widely. In their landmark paper, Baekeland and Lundwall[2] commented that "definitional inexplicitness has been the hallmark of most studies of dropping out of treatment."

Thomas Wickizer, Adam Atherly, Margaret Frederick, and Thomas Koepsell are with the Department of Health Services, University of Washington, Seattle. Adam Atherly is also with the Department of Economics, University of Washington. Thomas Koepsell is also with the Department of Epidemiology, University of Washington. Charles Maynard is with the Department of Medicine, School of Medicine, University of Washington. Antoinette Krupski and Kenneth Stark are with the Washington State Department of Social and Health Services, Division of Alcohol and Substance Abuse, Olympia.

Requests for reprints should be sent to Thomas Wickizer, PhD, Department of Health Services, SC-37, University of Washington, Seattle, WA 98195.

This paper was accepted July 26, 1993.

The lack of conceptual clarity as to who should be categorized as a dropout has fueled inconsistencies. For example, patients who make an appointment but do not keep it are sometimes, but not always, classified as dropouts. Similarly, patients leaving against medical advice but returning later to complete treatment can be classified as dropouts and then new entrants or as completers. Patients asked to leave treatment may be categorized as dropouts or simply excluded from the sample. As a result of these inconsistencies, studies have tended to define program completion in terms of criteria related to length of time in treatment[3] or achievement of treatment plan goals,[4,5] and some studies have simply left treatment completion undefined.[6]

Another problem affecting the reliability of findings reported in the literature is small sample size. Studies based on samples of fewer than 200 subjects,[7–13] or even fewer than 100 subjects[14–17] are not unusual. Even studies with large samples are often limited to a single treatment setting, prompting concern about external validity.[5,18] Many studies exhibit both problems.[19–21] Thus, considerable caution is in order when reviewing findings of studies conducted to assess completion rates.

Factors Associated with Treatment Completion

Dropout rates reported by inpatient drug studies have ranged from 19%[7] to 63%[22]; outpatient drug studies usually report higher dropout rates, with proportions above 70% being the norm.[4,19] Alcohol studies have followed the same pattern, with inpatient dropout rates ranging from 17.4%[18] to 74%[14] and outpatient rates often exceeding 70%.[2,8,15,20]

In general, the drug and alcohol treatment literature has examined the impact on completion of one or more of three characteristics: psychological, demographic, and program related. Psychological studies have found differences between dropouts and completers in terms of several factors, including impulse control, interpersonal difficulties, authority conflicts, motivation to succeed in treatment, and self-concept.[3,6,10,14,21,23,24] However, researchers have questioned the effectiveness of examining psychological factors.[25]

Studies using demographic variables have found that completers tend to be older, married with dependents at home, and wealthier and tend to have health problems related to drinking (such as cirrhosis).[6,14,26,27] Men tend to have higher completion rates than women,[27] but this result is weakened both by the lack of women in most studies[18] and by the greater percentage of men coerced into treatment. Ethnicity has not had a consistent impact, but studies have shown that the ethnic group in the majority within the treatment setting tends to have higher completion rates.[28]

Program variables have also been examined with mixed results. Studies have shown that dropout rates can be reduced by changing the therapeutic atmosphere of the program.[6,29–31] For example, completion rates have been linked to treatment staff vacation schedules[5,32] and to therapy group size.[33] Despite these findings, the long-term impact of program factors on outcomes has been questioned.[11]

Methods

The Substance Abuse Monitoring System database contains detailed information on client characteristics and services provided to individuals whose treatment is supported by state funds. Given the size of this database (upwards of a million records), it was necessary to sample records selectively. The study population was limited to clients discharged during the last quarter of 1990 from the following adult treatment modalities: intensive inpatient treatment, long-term inpatient treatment, intensive outpatient treatment, standard outpatient treatment, and recovery house (a therapeutic setting that provides transitional care for clients moving from inpatient to outpatient treatment). Three types of youth programs were also included: intensive inpatient, intensive outpatient, and standard outpatient. Methadone maintenance and detoxification programs were excluded.

Initial data were obtained on 6682 client records, representing all discharges from the above modalities during the fourth quarter of 1990. These records were loaded into the Rbase database management system and examined for incomplete data as well as out-of-range numeric and alphabetic codes. Of the 6682 records, only 3 had to be discarded because of excessive missing data or nonsensical information. Discharge information indicated that 10 clients had died in treatment and that 110 clients had been inappropriately admitted to treatment. These 120 records were dropped, leaving 6559 records available for analysis. However, 90% (5827) of these records represented discharges in three adult treatment modalities (intensive inpatient, intensive outpatient, and standard outpatient treatment) and one youth treatment modality (standard outpatient treatment). To simplify this study, it was decided to limit the analysis to the 5827 records representing clients discharged from these four modalities.

Program completion status was defined in terms of the discharge categories contained in the system database. These categories included (1) against advice, (2) rule violation, (3) client died, (4) inappropriate admit, (5) other, (6) transfer, (7) default, and (8) completed treatment. All clients with the completed treatment discharge code were defined as program completers. Clients transferred to other programs meeting certain time-in-treatment criteria were also considered to have completed treatment. Clients with the against advice, default, and rule violation discharge codes, as well as transferred clients not meeting the time-in-treatment criteria, were defined as noncompleters.

The judgment of whether a client completed treatment was based primarily on criteria related to length of treatment and secondarily on treatment plan factors. Clients in intensive inpatient programs were expected to remain in treatment from 21 to 28 days. Clients in intensive outpatient programs were expected to stay in treatment at least 1 month but could stay up to 3 months. Furthermore, these clients were to receive counseling and other treatment services at least three times per week for the first month. Standard outpatient treatment usually involved less frequent therapy; the only requirement was that the client be seen once per month unless he or she was under court supervision or some other binding order. No explicitly defined minimum target length of time or maximum time limit was established for standard outpatient treatment. Treatment completion for this modality was defined primarily in terms of progress toward meeting treatment plan objectives.

The Substance Abuse Monitoring System database contains mainly client-related variables. The major variables included in this analysis were age, gender, years of education, ethnic group, usual living arrangement (family vs other), marital status, number of children in household, employment status, referral source (courts, assessment center, other), primary substance abused, income, public assistance status, and existence of a secondary or tertiary substance abuse problem.

Two primary methods of analysis were used to assess program completion rates: cross-tabulation analysis and stepwise logistic regression. Cross-tabulation analysis provided comparative information on program completion rates for client and program variables. Stepwise logistic regression analysis identified variables that were significantly and independently associated with treatment completion ($P < .05$) and provided odds ratios (ORs) and 95% confidence intervals (CIs) for these variables. Separate analyses were conducted for adult drug and adult alcohol treatment programs. Youth programs, which do not distinguish between drug and alcohol clients, were analyzed separately as a single category.

Results

Table 1 presents descriptive data on the characteristics of the adult study population. As shown, the modality with the greatest number of discharges (2082) was standard outpatient alcohol treatment, followed by intensive inpatient alcohol treatment (973). Overall, the three alcohol treatment modalities accounted for almost 70% of discharges of adult clients. Women accounted for approximately 30% of the discharges from alcohol treatment programs and for more than 40% of the discharges from drug treatment programs. Approximately 65% of the adult clients were between 20 and 40 years of age, and more than 70% were White. Blacks represented roughly 15% to 20% of the discharges from drug treatment programs. Native Americans accounted for 16% of the discharges from inpatient alcohol programs and 8% of the discharges from outpatient alcohol programs. Hispanic clients represented a smaller proportion of the study population, roughly 2% to 7% depending on the modality. Most of the adult clients discharged had at least some high school education, and about 15% had attended school beyond this level.

The nature of drug and alcohol problems varied across modalities. A small percentage of clients in adult alcohol programs abused substances other than alcohol, reflecting the mixed nature of addiction developed by some clients (nearly all of these clients had alcohol abuse as a secondary problem). Cocaine abuse was the primary problem for almost half of the clients in adult drug treatment programs, while marijuana abuse was the primary problem for roughly 20% to 25% of the clients. Clients in drug treatment programs appeared to abuse multiple substances more often than clients in alcohol treatment programs. For example, more than 85% of the clients in drug treatment programs had a secondary drug problem, and roughly 55% to 62% had a tertiary drug problem. In contrast, 47% to 66% of alcohol treatment clients had a secondary drug problem, and less than 40% had a tertiary drug problem. As Table 1 shows, clients came into treatment through different avenues. Roughly 90% of the clients in inpatient programs were referred through a central assessment center, while 44% and 62% of the clients entering standard outpatient and intensive outpatient alco-

TABLE 1—Descriptive Information (%) on Clients in Adult Treatment Programs

	Alcohol Programs			Drug Programs		
	Intensive Inpatient (n = 973)	Standard Outpatient (n = 2082)	Intensive Outpatient (n = 832)	Intensive Inpatient (n = 376)	Standard Outpatient (n = 863)	Intensive Outpatient (n = 266)
Female	29	29	28	48	43	47
Age, y						
<20	6	7	5	4	17	11
20–29	38	33	37	57	42	45
30–39	37	34	36	31	33	34
40–49	14	17	14	7	8	10
≥50	5	9	8	1	1	<1
Ethnic group						
Asian	<1	1	<1	<1	1	1
Black	8	7	7	20	15	23
White	72	78	78	72	74	70
Hispanic	4	6	7	4	4	2
Native American	16	9	8	4	6	4
Education, y						
<9	8	8	8	7	9	8
9–12	78	74	72	80	73	74
13–16	14	17	20	13	17	17
>16	1	1	<1	1	<1	<1
Monthly income						
<$100	71	21	15	66	21	29
$100–$499	20	36	29	21	51	37
$500–$899	7	21	26	10	15	24
≥$900	2	23	30	3	14	9
Married	18	25	26	21	17	20
Living with family	45	49	50	42	50	46
One or more children	28	33	42	30	35	40
Employment status						
Full time	1	26	34	1	10	13
Part time/temporary	1	10	11	3	7	9
In school	1	5	3	1	15	8
None	97	59	52	96	67	70
Public assistance	23	27	22	30	38	37
First year of addiction						
Prior to 1950	2	4	3	...	<1	...
1950–1959	7	9	8	2	2	<1
1960–1969	24	23	24	6	10	14
1970–1979	25	24	22	36	32	34
1980 and later	25	24	22	56	56	52
Primary drug problem						
Alcohol	91	97	99	7	8	12
Heroin	<1	<1	<1	18	12	8
Amphetamines	<1	<1	...	10	8	11
Cocaine	4	1	1	45	40	46
Marijuana	3	1	<1	19	26	21
Other	<1	<1	...	2	7	2
Secondary drug problem	66	47	49	89	87	90
Tertiary drug problem	38	25	23	62	55	56
Referral type						
Assessment center	92	25	13	92	31	23
Courts	<1	44	62	<1	18	25
Other	7	31	25	7	51	52

TABLE 2—Descriptive Information on Clients in Outpatient Youth Treatment Programs (n = 435)

	%
Female	44
Age, y	
10–14	28
15–19	72
Ethnic group	
Asian	1
Black	5
White	85
Hispanic	5
Native American	4
Education, y	
≤8	37
9–12	62
Employment status	
Full time	1
None	18
Part time/temporary	4
In school	76
First year of addiction	
1970–1979	5
1980 and later	95
Primary drug problem	
Alcohol	60
Cocaine	2
Marijuana	28
Other/unknown	9
Secondary drug problem	66
Tertiary drug problem	26
Referral type	
Courts	23
Other	77

hol treatment programs, respectively, were referred through the courts.

Table 2 presents descriptive information on the 435 clients discharged from youth outpatient treatment programs during the study period. As shown, 44% of these clients were female, 28% were under 15 years of age, 85% were White, and 76% were attending school. Sixty percent of the youth clients received treatment for alcohol abuse problems; 30% received treatment for marijuana abuse. Sixty-six percent had a secondary drug problem, 26% had a tertiary drug problem, and 23% entered treatment as a result of a court order.

Table 3 presents information on treatment completion for adult clients and shows the results of the cross-tabulation analysis for selected variables. The first noteworthy items are the overall completion rates. As shown, treatment modality was associated with program completion ($P < .001$). Completion rates for intensive inpatient, intensive outpatient, and standard outpatient adult alcohol treatment programs were 75%, 23%, and 40%, respectively. For adult drug treatment programs, completion rates for these same three modalities were 71%, 19%, and 33%, respectively.

Variables found to be associated with completion of adult alcohol or drug treatment programs included age, education, ethnic group, year of addiction, public assistance, and referral source. In general, older clients and clients with more years of education had higher completion rates, but the differences were not always statistically significant. Eighty percent of White clients completed intensive inpatient alcohol treatment; this rate was higher than that for Blacks (71%), Hispanics (73%), and Native Americans (60%). However, the opposite was true for intensive inpatient drug treatment programs, in which 67% of White clients completed treatment as compared with 83% of Black clients, 93% of Hispanic clients, and 77% of Native American clients. Hispanic clients in outpatient drug programs had higher completion rates than clients of other ethnic groups. In almost all cases, clients whose addictions started earlier were more likely to complete treatment.

Finally, completion rates varied by referral source. In particular, adult clients entering treatment through an assessment center that provided centralized diagnostic, treatment planning, and referral services were more likely to complete treatment than clients entering treatment through other avenues, such as court referral. For example, 77% of clients in intensive inpatient alcohol treatment programs referred through an assessment center completed treatment, as compared with 58% of clients referred by the courts ($P = .003$). These results must be considered with some caution, because the many comparisons increase the likelihood of detecting a significant association.

Although analyzed, completion rates for youth clients are not shown in Table 3 to simplify the presentation of results. The overall completion rate for clients discharged from youth outpatient programs was 27%. While there were differences in completion rate among variables, in only one case did these differences achieve statistical significance. Youth clients whose primary drug problem was alcohol or marijuana abuse had higher completion rates than did other clients ($P = .09$).

Finally, stepwise logistic regression was used to identify the variables independently associated with treatment completion (Table 4). The variables most consistently related to treatment completion were age, education, and referral source. Older clients and clients with more years of education were more likely to complete inpatient as well as outpatient treatment. Assessment center screening was a strong independent predictor of treatment completion. Clients discharged from intensive outpatient drug treatment programs screened through an assessment center were almost five times as likely to complete treatment as were other clients.

Ethnicity showed a mixed pattern of relationships. For Native American clients in inpatient alcohol programs, the odds of completion were only about half as great as for other clients. Similarly, White clients were less likely to complete inpatient drug treatment but more likely to complete standard outpatient drug treatment. The odds of Blacks completing intensive outpatient drug treatment were only 15% as great as for other clients. Finally, clients with a secondary drug problem were less likely to complete inpatient alcohol treatment than clients without a secondary problem.

Regression analysis was also performed on data for youth clients. Although not shown in Table 4, two variables were found to be related to the odds of youth clients completing treatment: monthly household income (OR = 1.03 per $100 increase, 95% CI = 1.01, 1.05) and number of children in household (OR = 1.17, 95% CI = 1.03, 1.31).

Discussion

Completion rates were found to be associated with a number of client variables and with treatment modality. The notable finding in regard to treatment modality was the substantially higher completion rates observed for inpatient programs relative to outpatient programs, a finding that has been widely reported* but not well understood. The limited data captured by the Substance Abuse Monitoring System precluded a thorough examination of the many possible reasons for these differences. As Table 1 shows, there were few meaningful differences in client characteristics between the inpatient and outpatient groups studied.

The differences may be related to unmeasured factors associated with programs' therapeutic milieu or treatment setting. Strong clinical norms and peer expectations exist in inpatient substance abuse programs

*References 2, 4, 7, 8, 14, 15, 17, 18–20, 31.

TABLE 3—Completion Rates for Adult Clients in Alcohol and Drug Abuse Treatment Programs, by Selected Variables

	Alcohol Treatment Programs						Drug Treatment Programs					
	Intensive Inpatient Treatment (n = 973), %	P	Intensive Outpatient Treatment (n = 832), %	P	Standard Outpatient Treatment (n = 2082), %	P	Intensive Inpatient Treatment (n = 376), %	P	Intensive Outpatient Treatment (n = 266), %	P	Standard Outpatient Treatment (n = 863), %	P
Overall completion rate	75		23		40	<.001	71		19		33	<.001
Gender		NS		NS		NS		NS		.020		NS
Female	75		20		35		69		13		34	
Male	76		24		34		74		24		32	
Age, y		.051		NS		<.001		.011		NS		NS
<20	66		21		18		62		7		38	
20–29	72		20		31		66		19		29	
30–39	79		24		37		78		21		37	
40–49	80		25		36		88		27		30	
≥50	77		23		41		100				36	
Ethnic group		<.001		NS		NS		.008		.001		.031
Asian	67		33		27		100				33	
Black	71		26		30		83		3		29	
White	80		23		35		67		24		35	
Hispanic	73		12		34		93		40		36	
Native American	60		23		31		77		9		15	
Education, y		.020		NS		NS		NS		NS		NS
<9	62		15		31		72		5		25	
9–12	76		22		32		71		19		34	
13–16	80		25		41		74		25		32	
>16	90		67		42		100		50		50	
Employment status		.121		NS		.005		NS		NS		NS
Full time	77		25		39		60		27		25	
Part time/temporary	80		18		26		90		21		32	
In school	33		13		30		100		9		39	
None	76		22		33		71		18		39	
Public assistance		NS		NS		NS		NS		NS		NS
Yes	74		18		31		75		13		28	
No	76		24		35		70		22		36	
First year of addiction		.086		.024		.012		NS		NS		NS
Prior to 1950	78		36		47						50	
1950–1959	76		18		35		83				31	
1960–1969	80		27		36		73		32		32	
1970–1979	77		23		34		72		25		32	
1980 and later	70		16		29		70		11		34	
Primary drug problem								NS		NS		NS
Alcohol							84		29		24	
Heroin							75		21		36	
Amphetamines							58		14		45	
Cocaine							70		14		30	
Marijuana							73		25		26	
Others							100		50		72	
Referral type		.003		NS		<.001		.090		<.001		<.001
Assessment center	77		26		44		73		38		45	
Courts	60		23		35		33		22		31	
Other	58		21		25		57		9		26	

that may make it more difficult for clients to abandon treatment prior to completion. Also, not completing treatment involves two very different actions depending on whether the individual is an inpatient or outpatient client. A client receiving inpatient treatment must physically leave the treatment setting in order not to complete treatment, while a client receiving outpatient treatment merely has to fail to show up for his or her scheduled therapy.

Unmeasured factors related to source of payment or cost sharing may account for some of the differences in completion rates between the inpatient and outpatient study groups. The treatment of nearly all inpatient clients was supported directly by state funding. However, support for outpatient clients was provided indirectly through the counties, and some income-related copayment may have been required. If so, this requirement may have posed a financial barrier to completion and may explain why fewer outpatient clients completed treatment.

A limited number of variables were found to be significantly predictive of treatment completion in multivariate analyses. In general, older clients, clients with more education, clients with a longer history of substance abuse, and clients referred to treatment through a central referral and assessment center were more

TABLE 4—Stepwise Logistic Regression Models Showing Predictors of Treatment Completion, by Modality

	Adjusted Odds Ratio	95% Confidence Interval	P
Adult alcohol programs			
Intensive inpatient treatment (n = 969)			
Primary substance of abuse other than alcohol	0.47	0.29, 0.75	.002
Years of education	1.18	1.09, 1.28	<.001
Secondary drug use	0.52	0.38, 0.76	<.001
Native American	0.45	0.31, 0.66	<.001
Intensive outpatient treatment (n = 814)			
Years of education	1.11	1.03, 1.20	.010
Later first year of addiction	0.84	0.71, 0.99	.040
Standard outpatient treatment (n = 2026)			
Screened by assessment center	2.61	2.00, 3.39	.001
Court referral	1.53	1.20, 1.94	<.001
Age (per 10-year increase)	1.02	1.01, 1.03	<.001
Income	1.16	1.05, 1.29	.004
Adult drug programs			
Intensive inpatient treatment (n = 374)			
Age (per 10-year increase)	1.96	1.34, 2.86	.001
White	0.42	0.23, 0.75	.005
Intensive outpatient treatment (n = 261)			
Screened by assessment center	4.84	2.14,10.91	<.001
Court referral	2.63	1.10, 6.27	.029
Later first year of addiction	0.55	0.35, 0.85	.008
Black	0.15	0.03, 0.66	.012
Standard outpatient treatment (n = 837)			
Primary drug other than alcohol, marijuana, cocaine, heroin	3.15	2.09, 4.75	<.001
Screened by assessment center	2.43	1.76, 3.36	<.001
White	1.55	1.07, 2.24	.005
Years of education	1.09	1.01, 1.18	.041

Note. The following independent variables were included in the logistic regression analysis: age, gender, disability, marital status, employment status, receiving public assistance, eligible for health insurance, White ethnic group, Black ethnic group, Hispanic, Native American, pregnant, Alcohol and Drug Addiction Treatment and Support Act (ADATSA) as funding source, screened by assessment center, fee status, heroin as primary drug, alcohol as primary drug, cocaine as primary drug, marijuana as primary drug, secondary drug use, tertiary drug use, veteran, monthly income, education, number of children in household, first year of addiction, number in household, and living with family. A few records contained missing information and were therefore excluded from the regression analyses. The figures shown represent the numbers of observations used for the regressions.

likely to complete treatment, other things equal.

The findings regarding ethnic group and referral mode merit discussion. Whites were more likely to complete outpatient drug treatment than were members of other ethnic groups, but they were less likely to complete inpatient drug treatment. Native Americans were less likely to complete inpatient alcohol treatment than were clients of other ethnic groups, and Blacks were less likely to complete intensive outpatient drug treatment. These findings may suggest the importance of cultural fit between clients and treatment programs.[28] Ethnicity per se may not be as important as the fit between one's ethnic group and the norms and culture of the program.

This notion of client–program fit may extend to other areas as well, particularly the fit between the client's substance abuse problem and the treatment focus. The regression analysis indicated that clients in intensive inpatient alcohol programs whose primary problem involved abuse of a substance other than alcohol were only about half as likely to complete treatment as were clients whose primary problem involved alcohol abuse.

The fit between client and program has important clinical implications and may explain why referral mode, particularly referral through a centralized assessment center, was found to be an important independent predictor of treatment completion. Assessment centers provide centralized intake, screening, and referral of clients, which may lead to a better fit between the program and client than would otherwise occur, for example, when clients are referred to treatment through the courts.

How do these results compare with other findings reported in the literature? While comparisons must be made cautiously, it appears that the completion rates described here are somewhat higher than those reported by other studies.[2,14,18,31]

One important limitation of the present study was its focus on a single treatment episode. Clients with substance abuse problems often move in and out of treatment over time. What may be important for achieving abstinence or positive outcomes (e.g., employment) in the long run is the cumulative effect of multiple treatments over time rather than the fact of treatment completion for a single treatment episode. The results of the regression analysis indicated that clients with fewer years of addiction were less likely to complete intensive outpatient drug and alcohol treatment than were clients with a longer history of addiction. The latter clients probably had greater exposure to treatment over time, and the cumulative effects of this previous treatment may have increased the likelihood of their completing the current treatment episode. Unfortunately, it was not feasible within the scope of this study to extract data from the Substance Abuse Monitoring System that would have allowed us to link records over time to construct a variable that could reliably measure past treatment.

Treatment completion is believed to be important because it presumably has a positive effect on long-term outcomes. It was not feasible, in the present study, to perform long-term follow-up on clients. Thus, we do not know whether the clients in this study who completed treatment had better long-term outcomes than those who did not complete treatment.

In addition to analyzing treatment completion, this study had a secondary purpose: to demonstrate the potential of a statewide information system to provide reliable data for analysis. Only 3 of the more than 6000 records we obtained for the project proved unusable as a result of poor data, a remarkable fact. Furthermore, an independent study of youth treatment programs, undertaken at the same time as our study and involving actual reviews of client files, generated completion-rate estimates virtually identical to those reported here.

Our experience suggests that state information systems can provide a rich

source of data, yielding information that is sufficiently complete and timely to provide a basis for meaningful analysis. Yet the capabilities of many state systems appear to be rather limited.[34] A critical need for improving understanding of treatment effectiveness through analysis of information contained in state information systems is the ability to track clients over time and to link treatment episodes. Building the capacity to do so, even on random samples of clients, would greatly enhance the capability of these information systems to provide meaningful data.

From a health policy perspective, the question of the effectiveness of substance abuse treatment has taken on added importance in the context of health care reform with discussion of expanding health insurance coverage for mental health and substance abuse treatment. On the private side, the cost of substance abuse treatment has risen dramatically in recent years and is now one of the fastest growing expenditure categories, a fact that has prompted employers and insurers to scrutinize treatment much more closely and to question its value. The need for information systems that can provide reliable data to support and enhance planning and evaluation of substance abuse treatment services will doubtlessly grow in the future.

Further study will be needed to better understand the factors predictive of program completion. This study suggested the notion that fit between the client and program may be a key factor explaining why some clients complete treatment and others drop out. In this regard, there is a need for careful systematic assessment of how different referral mechanisms may influence client–program fit. More generally, there is a need to better understand how program factors, which may be amenable to change through health policy or management initiatives, may affect completion rates as well as long-term outcomes. □

Acknowledgments

Financial support for this research was provided by the Washington State Department of Social and Health Services, Division of Alcohol and Substance Abuse.

References

1. *National Drug and Alcoholism Treatment Unit Survey (NDATUS) 1989: Main Findings Report.* Rockville, Md: National Institute on Drug Abuse and National Institute on Alcohol Abuse and Alcoholism; 1991.
2. Backeland F, Lundwall L. Dropping out of treatment: a critical review. *Psychological Bull.* 1975;82:738–783.
3. Wilkinson E, Prado W, Williams W, Schnadt F. Psychological test characteristics and length of stay in alcoholism treatment. *Q J Stud Alcohol.* 1971;32:60–65.
4. Steer R. Retention in drug-free counseling. *Int J Addict.* 1983;18:1109–1114.
5. Fink E, Rudden S, Longabaugh R, McCrady B, Stout R. Adherence in a behavioral alcohol treatment program. *Int J Addict.* 1984;19:709–719.
6. Linn M. Attrition of older alcoholics from treatment. *Int J Addict Disord.* 1978;3:437–447.
7. Craig R, Olsen R. Differences in psychological need hierarchy between program completers and dropouts from a drug abuse treatment program. *Am J Addict.* 1990;25:1–26.
8. Rees D, Beech H, Hore B. Some factors associated with compliance in the treatment of alcoholism. *Alcohol Alcohol.* 1984;19:303–307.
9. Siddall J, Conway G. Interactional variables associated with retention and success in residential drug treatment. *Int J Addict.* 1988;23:1241–1254.
10. Schofield L. Internal-external control and withdrawal AMA from an alcohol rehabilitation program. *J Clin Psychol.* 1978;34:571–573.
11. Brizer D, Maslansky R, Galanter M. Treatment retention of patients referred by public assistance to an alcoholism clinic. *Am J Drug Alcohol Abuse.* 1990;16:259–264.
12. Feigelman W. Day-care treatment for multiple drug abusing adolescents: social factors linked with completing treatment. *J Psychoactive Drugs.* 1987;19:335–344.
13. Allan C. Seeking help for drinking problems from a community-based voluntary agency. Patterns of compliance amongst men and women. *Br J Addict.* 1987;82:1143–1147.
14. Jones JW. Predicting patients' withdrawal against medical advice from an alcoholism treatment center. *Psychological Rep.* 1985;57:991–994.
15. Addenbrooke W, Rathod N. Relationship between waiting time and retention in treatment amongst substance abusers. *Drug Alcohol Depend.* 1990;26:255–264.
16. Roffe M. Predictive correlates of treatment program completion in a sample of male alcoholics. *Int J Addict.* 1981;16:849–857.
17. Means L, Small M, Capone D, et al. Client demographics and outcome in outpatient cocaine treatment. *Int J Addict.* 1989;24:765–783.
18. Brewer L, Zawadski M, Lincoln R. Characteristics of alcoholics and codependents who did and did not complete treatment. *Int J Addict.* 1990;25:653–663.
19. Stark M, Campbell B. Personality, drug use, and early attrition from substance abuse treatment. *Am J Alcohol Drug Abuse.* 1988;14:475–485.
20. Leigh G, Ogborne J, Cleland P. Factors associated with patient dropout from an outpatient alcoholism treatment service. *J Stud Alcohol.* 1984;45:359–362.
21. Huber N, Danahy S. Use of the MMPI in predicting completion and evaluating changes in a long-term alcoholism treatment program. *J Stud Alcohol.* 1975;36:1230–1237.
22. Harris R, Linn M, Pratt T. A comparison of dropouts and disciplinary discharges from a therapeutic community. *Int J Addict.* 1980;15:749–756.
23. Krasnoff A. Differences between alcoholics who complete or withdraw from treatment. *J Stud Alcohol.* 1976;37:1666–1771.
24. Wilson E, Whelan W. Anxiety as a factor in continuance and dropout in treatment. *Int J Addict.* 1983;18:577–582.
25. Krasnoff A. Failure of MMPI scales to predict treatment completion. *J Stud Alcohol.* 1977;38:1440–1442.
26. Hahn J, King K. Client and environmental correlates of patient attrition from an inpatient alcoholism treatment center. *J Drug Educ.* 1982;12:75–86.
27. Sansone J. Retention patterns in a therapeutic community for the treatment of drug abuse. *Int J Addict.* 1980;25:1–26.
28. Brown B, Joe G, Thompson P. Minority group status and treatment retention. *Int J Addict.* 1985;20:319–335.
29. Moise R, Reed B, Conell C. Women in drug abuse treatment programs: factors that influence retention at very early and later stages in two treatment modalities. A summary. *Int J Addict.* 1981;16:1295–1300.
30. Verinis JS. Ward atmosphere as a factor in irregular discharge from an alcohol rehabilitation unit. *Int J Addict.* 1983;18:895–899.
31. Craig R. Reducing the treatment drop out rate in drug abuse programs. *J Subst Abuse Treatment.* 1985;2:209–219.
32. Brown W, Twemlow S. Staff absence as a factor in the patient dropout rate in alcoholism treatment programs. *Hosp Community Psychiatry.* 1978;29:361–367.
33. Schroeder D, Bowen W, Twemlow S. Factors related to patient attrition from alcoholism treatment programs. *Int J Addict.* 1982;17:463–472.
34. *National Survey on Select Treatment and Prevention Issues.* Olympia, Wash: Division of Alcohol and Substance Abuse, Washington State Dept of Social and Health Services; 1991.

Ethnicity and the Use of Outpatient Mental Health Services in a National Insured Population

Deborah K. Padgett, PhD, Cathleen Patrick, PhD, Barbara J. Burns, PhD, and Herbert J. Schlesinger, PhD

ABSTRACT

Objectives. Factors affecting ethnic differences in the use of outpatient mental health services are analyzed in an insured, nonpoor population to determine if lower use by Blacks and Hispanics persists when socioeconomic and other factors are controlled.

Methods. To identify significant predictors of the probability and amount of use, insurance claims data for a population of 1.2 million federal employees insured by Blue Cross/Blue Shield in 1983 were analyzed with the Andersen and Newman model of health service utilization. Logistic and ordinary least squares regression models were estimated for each ethnic group.

Results. Blacks and Hispanics had lower probabilities and amounts of use when compared with Whites after controlling for a number of variables.

Conclusions. Since ethnic differences in the use of outpatient mental health services exist even in an insured, nonpoor population, factors other than lower socioeconomic status or insurance coverage—for example, cultural or attitudinal factors and service system barriers—are likely responsible. Such findings have policy implications in the current climate of health care reform to increase access to care for the underserved. (*Am J Public Health.* 1994;84: 222–226)

Introduction

National priorities supporting mental health research on ethnic minorities were set forth in the 1978 report of the President's Commission on Mental Health.[1] However, most studies of ethnic use of mental health services have been confined to public sector institutions such as state and federal psychiatric hospitals and community mental health clinics; use of private sector mental health services is rarely captured.

While information on all forms of mental health services use by ethnic groups is critical to establishing sound policies in financing and services delivery, relatively little is known about how members of ethnic groups use outpatient mental health services, especially when comprehensive insurance coverage is available to them. A review of the literature on ethnicity and the use of outpatient mental health services reveals two consistent patterns. First, all ethnic/racial groups (including Whites) underuse mental health services relative to need, as indicated by the Epidemiologic Catchment Area studies sponsored by the National Institute of Mental Health in the early 1980s,[2–6] which were instrumental in documenting gaps between need for and use of mental health services in the general US population.[3,5,6] Second, findings from these and other studies have consistently shown that the gap between need and use is greatest for members of ethnic minority groups.[7–11] Thus, rates of outpatient mental health services use tend to be lower among Blacks and Hispanics than among Whites.

Ethnic differences in the amount of mental health services used have also been noted, although findings are not always consistent. While Blacks, Asian Americans, and Hispanics generally make fewer outpatient visits than Whites,[8,12–15] Wells et al.[10] and Hu et al.[16] report no significant ethnic differences in the amount of use once treatment has begun.

Some researchers have noted that interethnic differences in mental health services use are the result of socioeconomic differences and are minimized when these factors are controlled.[10,15,17] However, little is known about how factors such as insurance coverage and service availability affect the likelihood of seeking outpatient mental health treatment among different ethnic groups and to what extent the influence of such factors differs across ethnic groups.

The first goal of this study is to assess factors influencing the use of outpatient mental health services among ethnic groups in an insured, nonpoor population. Our database of more than 1.2 million federal employees and their family members insured by Blue Cross/Blue Shield in 1983 offered a rare opportunity to examine and compare service use by Whites, Blacks, and Hispanics in a large national population.

The second goal is to discuss the policy implications of this study's findings with regard to the financing and delivery of mental health services to ethnic minority groups in the 1990s. The past decade has brought enormous increases in health

Deborah K. Padgett is with New York University's School of Social Work, New York, NY. Cathleen Patrick and Herbert J. Schlesinger are with the Department of Psychology at the New School for Social Research, New York, NY. Barbara J. Burns is with the Department of Psychiatry, Duke University Medical Center, Durham, NC.

Requests for reprints should be sent to Deborah K. Padgett, PhD, School of Social Work, New York University, 3 Washington Square N, New York, NY 10003.

care costs along with renewed calls for national health insurance.[18] We argue that this study's findings can provide insight into ethnic patterns of use under a system of national health insurance.

Methods

Database

The study used insurance claims and related enrollment data from federal employees and their family members who were insured by the Blue Cross and Blue Shield Association's Federal Employees Plan in 1983. For outpatient mental health care, the high-option plan copayment was 30% with a $200 deductible and a 50-visit maximum, and the low-option plan copayment was 25% with a $250 deductible and a 25-visit maximum. The deductible included all medical claims, not just those for mental health treatment. It is likely that usage below the deductible is included in these data since Blue Cross/Blue Shield encouraged enrollees to file claims even when those claims did not exceed the deductible.

The study sample comprised all persons who had at least one outpatient mental health visit in 1983 and random samples of approximately 5000 nonusers selected within each ethnic group. In subsequent analyses, these samples were weighted to estimate the total number of persons who did not have outpatient mental health treatment. The preexisting structure of the database required that only families enrolled for 5 or more years were included.

Predictors of Probability and Amount of Use

Following the Andersen and Newman model of health service use,[19] three sets of independent variables were developed: predisposing, enabling, and need factors. Most of the predisposing variables are self-explanatory (e.g., age, sex, etc.). However, the rationale for using the percentage of the county that was Black, Hispanic, or White as a measure of ethnic congruity was as follows: Tweed et al.[20] found that persons residing in areas characterized by the numerical dominance of own-group members manifested lower levels of psychological distress when compared with persons living in racially "dissonant" areas. We reasoned that the level of ethnic congruence in a person's geographic area might also influence that person's decision to seek mental health care.

TABLE 1—Outpatient Mental Health Visits and Mean Total Visits (for Those with at Least One Visit) in 1983

	Blacks (n = 137 936)			Hispanics (n = 30 623)			Whites (n = 809 000)		
	% in Group	% with Visit	Mean No. of Visits	% in Group	% with Visit	Mean No. of Visits	% in Group	% with Visit	Mean No. of Visits
Sex									
Male	42.4	1.95	9.5	47.0	2.51	8.5	47.5	3.16	13.1
Female	57.6	2.48	9.9	53.0	3.00	8.7	52.5	4.12	12.9
Age, y									
0–17	17.3	1.68	9.6	21.5	1.88	9.7	15.5	3.08	12.7
18–45	29.1	3.47	11.5	24.9	3.78	11.6	24.9	5.82	16.7
46 or more	53.6	1.77	7.9	53.6	2.66	6.3	59.6	2.92	9.9
Region									
DC area	42.3	2.42	11.9	4.3	4.61	22.1	22.6	4.99	17.4
Northeast	9.8	1.72	11.9	3.1	2.35	15.4	17.9	3.48	14.3
North central	19.9	2.22	8.9	3.7	3.05	7.8	19.1	3.10	10.6
West	6.1	2.84	8.4	34.5	2.70	7.2	12.7	3.80	12.0
South	21.9	2.04	5.4	54.5	2.68	7.4	27.8	3.03	8.3
Option									
Low option	23.6	1.98	6.7	23.5	1.93	5.8	41.6	2.59	9.0
High option	76.4	2.34	10.6	76.5	3.03	9.2	58.4	4.42	14.6
Years of employee education									
<12 y	27.9	1.73	7.3	35.5	2.42	6.2	11.7	2.55	7.0
High school graduate	49.3	2.27	9.1	43.4	2.57	7.6	49.7	2.96	10.4
Some college	12.2	2.44	11.6	10.4	3.22	10.6	10.6	4.00	12.6
College graduate	10.6	3.27	13.8	10.7	3.96	14.6	28.0	5.40	17.0
Overall	100.00	2.26	9.7	100.00	2.77	8.6	100.00	3.66	13.0

Enabling factors included region, salary of the employee, and high- or low-option plan. Region of country was considered an enabling factor since the availability of mental health services varies considerably.[21] For analyses predicting the number of visits, the setting of the first visit and the type of mental health practitioners sought were included. As classified in the claims data, such practitioners included physicians, psychologists, and mental health workers who were clinical social workers and psychiatric nurses. Physician specialty was not available from the Federal Employees Plan data and may have included either psychiatrists or general practitioners. Two enabling variables from the Bureau of Health Professions Area Resource File provided county-level measures of the relative availability of outpatient mental health services: the percentage that was urban and the ratio of psychiatrists to physicians.

Although specific mental diagnoses were unavailable, several variables were useful as indices of need for mental health treatment: the individual's and the rest of the family's annual medical expenses, prior inpatient psychiatric treatment during the year by anyone else in the family, and the total number of mental health visits made by all other family members. For the amount-of-use analyses, whether the individual also had inpatient mental health treatment during the year was included as a dichotomous yes/no variable. (In analyses predicting any use of outpatient mental health services, the term "individual" refers to any member of the study population; in analyses predicting the amount of use, it refers to the user of outpatient mental health services.)

Dependent Variables

Following previous analyses of mental health services use,[22–24] probability of use and amount of use constituted the two dependent variables. The first was coded "1" if the individual had a visit claim coded "nervous and mental" during the year, and "0" otherwise. The second variable was the number of mental health visits made during the year given any use.

Data Analysis

Weighted logistic regression models were developed separately for each ethnic group to calculate odds ratios (ORs) and 95% confidence intervals (CIs) to predict the probability of use. For predicting the amount of use, ordinary least squares regression was used. Variables were en-

TABLE 2—Prediction of Probability of Having at Least One Outpatient Mental Health Visit in 1983

	Blacks		Hispanics		Whites	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Predisposing factors						
Females vs males[a]	1.28	0.94, 1.74	1.20	0.87, 1.65	1.32*	1.10, 1.58
Age 0–17 vs 46 or more[a]	0.78	0.58, 1.04	0.71*	0.52, 0.95	0.82*	0.69, 0.98
Age 18–45 vs 46 or more[a]	1.40*	1.19, 1.64	1.19*	1.00, 1.42	1.43*	1.30, 1.57
Education of employee	1.07*	1.00, 1.14	1.05	1.00, 1.11	1.11*	1.07, 1.14
Family size	1.04	0.86, 1.25	1.05	0.89, 1.25	0.98	0.88, 1.08
% ethnic in county[b]	1.00	0.99, 1.00	1.00	0.99, 1.00	1.00	0.99, 1.00
Enabling factors						
Northeast vs DC[a]	0.74	0.40, 1.35	0.50	0.16, 1.61	0.82	0.62, 1.09
North central vs DC[a]	0.93	0.60, 1.43	0.65	0.24, 1.77	0.70*	0.53, 0.93
West vs DC[a]	1.16	0.62, 2.17	0.65	0.31, 1.34	0.79	0.58, 1.07
South vs DC[a]	0.87	0.58, 1.32	0.68	0.31, 1.50	0.69*	0.54, 0.90
% urban in county	1.00	1.00, 1.01	1.01	1.00, 1.01	1.00*	1.00, 1.01
Salary[c]	1.04	0.86, 1.26	1.01	0.84, 1.22	1.05	0.97, 1.14
High vs low option[a]	1.33	0.91, 1.93	1.70*	1.10, 2.63	2.11*	1.72, 2.59
Psychiatrist–physician ratio[c]	1.08	0.64, 1.80	0.87	0.46, 1.64	1.16	0.89, 1.51
Need factors						
Medical dollars spent on individual[c]	1.01	1.00, 1.02	1.03*	1.01, 1.05	1.01*	1.00, 1.02
Medical dollars spent on others in family[c]	1.00	0.98, 1.03	1.02	0.99, 1.05	1.00	0.99, 1.02
Outpatient mental health visits by others in family[c]	1.30*	1.17, 1.44	1.37*	1.22, 1.53	1.27*	1.21, 1.33
Others in family, mental health inpatients (yes vs no)[a]	1.59	0.60, 4.22	2.10	0.83, 5.29	1.76	0.93, 3.33

[a]The first mentioned group was coded "1" and the second was coded "0."
[b]In the regression analysis for Blacks, this is the percentage of the county that is Black (and similarly, for Hispanics and Whites in their respective regression analyses).
[c]Odds ratios for the continuous variables used in the model and their corresponding unit increases are as follows: (1) salary in units of $10 000, (2) psychiatrist to physician ratio in units of 0.1, (3) total medical charges for the sampled individual and for the rest of the family in units of $1000, (4) outpatient mental health visits for others in the family in units of five visits.
*Variables were statistically significant at $P = .05$.

tered into the regression equation following hierarchical procedures recommended by Cohen and Cohen.[25]

Results

Characteristics of the Population

Men made up a slightly greater proportion of Whites and Hispanics than of Blacks (see Table 1). There were noticeable differences in regional distributions for the three groups. It is difficult to explain why three fourths of Blacks and Hispanics chose the high-option plan while only 58% of Whites chose this plan. Federal employee education was substantially higher for Whites.

Outpatient Mental Health Services Use

As shown in Table 1, Blacks had the lowest rate of outpatient mental health services use, followed by Hispanics and then Whites. Sociodemographic breakdowns resulted in similar patterns across the groups, with women, 18- to 45-year-olds, high-option enrollees, and the more educated having higher rates. For Whites and Hispanics, the highest rate of use was in the Washington, DC, area; for blacks, it was in the western states.

Hispanics averaged the fewest number of visits, followed by Blacks and then Whites (see Table 1). While there were few gender differences, 18- to 45-year-olds, those living in the Washington, DC, area and in the Northeast, the more educated, and high-option enrollees averaged more visits.

Predicting the Probability of a Mental Health Visit

Table 2 shows that, for each group, being aged 18 to 45 as compared with being older was statistically significant (for Blacks: OR = 1.40; CI = 1.19, 1.64; for Hispanics: OR = 1.19; CI = 1.00, 1.42; for Whites: OR = 1.43; CI = 1.30, 1.57), as were the number of mental health visits made by other family members (for Blacks: OR = 1.30; CI = 1.17, 1.44; for Hispanics: OR = 1.37; CI = 1.22, 1.53; for Whites: OR = 1.27; CI = 1.21, 1.33). For Blacks, the only other significant predictor was education (OR = 1.07; CI = 1.00, 1.14). Education was also a significant predictor for Whites (OR = 1.11; CI = 1.07, 1.14) but not for Hispanics (OR = 1.05; CI = 1.00, 1.11).

Additional significant predictors for Hispanics were being over age 46 compared with being under age 18, having a high-option plan, and having higher medical costs for the individual. For Whites, these same variables were significant, with the addition of female gender, living in the District of Columbia area compared with living in the north central or southern states (odds ratios less than 1 require reversing the coding for these variables), and percentage urban population in the county of residence.

Predicting the Number of Mental Health Visits

Results in Table 3 reveal that, while many of the significant variables are predictably associated with higher levels of use, other predictors are lesser known. For example, ethnic congruence was significant for all three groups, although not in the same direction: it was positively associated with the number of visits for Blacks, but it was negatively associated for Hispanics and Whites.

Provider type and setting were also significant for all groups, with more visits associated with visiting an office (vs a clinic or hospital) and with visiting a psychologist (vs a physician). Medical costs of the individual and of others in the family were generally inversely related to the number of visits. Use of inpatient mental health services was positively and significantly related to the amount of outpatient use for Hispanics and Whites but not for Blacks.

Do Ethnic Groups Differ after Adjustment for All Covariates?

In a test of the overall hypothesis of whether statistically significant ethnic group differences remained after adjusting for all covariates in the study (comparing all three groups simultaneously), Whites had 1.7 times greater odds of making a visit compared with Blacks and Hispanics ($P < .004$). There was no significant difference between Black and Hispanic odds of making a visit. In the ordinary least squares regression model, Whites were

estimated to make 2.64 more mental health visits during the year than Hispanics and Blacks ($P < .001$). There was no significant difference found between Hispanic and Black users.

Discussion

This study has demonstrated that ethnic differences in the use of outpatient mental health services persist even in an insured population after a number of potentially confounding factors are controlled. The significance of enrollment in the high-option plan for Hispanics and Whites (but not for Blacks) is an indication that use of outpatient mental health services is responsive to insurance coverage in these groups. Although it is difficult to explain why plan option is not significant for Blacks in predicting the probability of use, it is easier to see how more generous coverage is significant for all groups in predicting higher amounts of use for those persons in treatment.

We also note that significantly higher levels of use are associated for all three groups with visiting an office-based practitioner and with visiting a psychologist. This effect of visiting a psychologist was also reported by Taube et al.[24]

While few of the indices of need proved to be influential, use of outpatient mental health services by other members of the family and higher medical costs of the individual were significant. These variables might be seen as indicating the influence of stress within the family and of physical comorbidity for the person seeking mental health treatment.

The cumulative results of this and previous studies point to a clear pattern of lower use by Blacks and Hispanics compared with Whites. While this phenomenon is more noticeable among the ethnic minority poor,[26] it is discernible even in this study's nonpoor population for whom many socioeconomic and other differences are minimized or controlled. We also note that the influence of differential acculturation is minimized in this population since unacculturated and/or non-English-speaking persons are unlikely to be included.

The reader is cautioned that these analyses captured usage during a 1-year period. Persons who used outpatient mental health services may have begun treatment prior to January 1, 1983, or continued treatment after the end of the year. Thus, our database restricted our analyses to annual usage rather than to episodes of treatment. We also acknowledge that the absence of diagnostic data limited our understanding of the nature and severity of the mental problems. However, some researchers have argued that diagnostic data may be more suspect in cross-ethnic comparisons.[27] Insurance claims data do have an advantage over self-reported data in studies of ethnic groups, in which underreporting of embarrassing or socially stigmatizing events such as visiting a mental health specialist has been noted.[27]

Given the enormous changes that have taken place in the delivery of mental health services since the early 1980s, this study's findings may at first glance appear to be of more historical value than current interest. Interestingly, there are no available sources of national data of more recent vintage. The most recent study allowing national estimates of rates and amounts of use—the National Medical Expenditure Survey—took place in 1987, but usage data were not yet available as of this writing.

However, we would argue that this study's findings are applicable to the current and future situation in the United States. Concerns about spiraling health care costs and the rapidly growing number of uninsured persons have led to renewed calls for national health insurance,[18,28] and several of the comprehensive proposals resemble the Federal Employee Plan's benefit structure.[18] Under national health insurance, low-income Blacks and Hispanics may gain the same coverage and access to mental health care as their employed counterparts had in the federal plan. As noted by Weil,[18] studies using resources such as the federal plan's database might provide

TABLE 3—Prediction of Number of Outpatient Mental Health Visits Made in 1983

	Blacks (n = 3106)		Hispanics (n = 849)		Whites (n = 8889)	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Predisposing factors						
Females vs males[a]	0.40	0.48	0.26	0.80	−0.28	0.30
Age 0–17 vs 46 or more[a]	−0.07	0.46	0.55	0.73	−0.48	0.29
Age 18–45 vs 46 or more[a]	1.79**	0.24	2.65**	0.43	3.41**	0.16
Education of employee	0.73**	0.09	0.73**	0.12	1.05**	0.05
Family size	0.01	0.29	−0.58	0.43	−0.29	0.16
% ethnic in county[b]	0.04**	0.01	−0.05**	0.02	−0.02**	0.01
Enabling factors						
Northeast vs DC[a]	−0.07	0.91	−7.84**	2.64	−2.28**	0.43
North central vs DC[a]	−2.63**	0.65	−15.35**	2.25	−5.60**	0.44
West vs DC[a]	−3.16**	0.94	−14.05**	1.62	−4.64**	0.47
South vs DC[a]	−6.32**	0.62	−13.64**	1.80	−7.42**	0.40
% urban in county	0.01	0.01	0.09**	0.02	0.04**	0.00
Salary[c]	0.94**	0.27	0.74	0.39	0.24*	0.12
High vs low option[a]	4.22**	0.55	3.32**	0.97	5.76**	0.30
Psychiatrist–physician ratio[c]	1.47	0.76	4.02*	1.59	2.73**	0.40
Office vs hospital	6.89**	0.44	3.70**	0.95	5.84**	0.41
Physician vs psychologist	−2.41**	0.31	−2.92**	0.50	−1.45**	0.16
Physician/psychologist vs other	−1.49**	0.54	0.82	1.09	−1.40**	0.34
Need factors						
Medical dollars spent on individual[c]	−0.09**	0.03	−0.12*	0.05	−0.11**	0.02
Medical dollars spent on others in family[c]	0.00	0.04	−0.06	0.06	−0.05*	0.02
Outpatient mental health visits by others in family[c]	1.26**	0.12	0.82**	0.13	0.93**	0.05
Others in family, mental health inpatients (yes vs no)[a]	−1.49	1.35	−2.45	1.73	−3.14**	0.84
Individual a mental health inpatient (yes vs no)[a]	0.61	0.53	1.76*	0.87	1.70**	0.40

[a]The first mentioned group was coded "1" and the second was coded "0."
[b]In the regression analysis for Blacks, this is the percentage of the county that is Black (and similarly, for Hispanics and Whites in their respective regression analyses).
[c]Regression weights for the continuous variables used in the model and their corresponding unit increases are as follows: (1) salary in units of $10 000, (2) psychiatrist to physician ratio in units of 0.1, (3) total medical charges for the sampled individual and for the rest of the family in units of $1000, (4) outpatient mental health visits for others in the family in units of five visits.
*Variables were statistically significant at $P = .05$.
**Variables were statistically significant at $P = .01$.

insight into how various ethnic groups use mental health services under a national plan. We acknowledge that Whites, Blacks, and Hispanics in the Federal Employee Plan population are not fully representative of their uninsured counterparts in the general US population, where the prevalence of severe mental disorders is reported to be higher[4]; differences in usage may be expected owing to a greater need associated with the lower socioeconomic status experienced by ethnic minorities. However, this study's primary finding that Whites use more mental health services even in an insured population points to the possibility that ethnic differences may persist even under national health insurance.

Clearly, the use of outpatient mental health services is governed by a number of factors related to the individual and the service delivery system. Many researchers suggest that cultural or attitudinal factors play a strong role in lower use by Blacks and Hispanics.[9,10,26,29,30] Such factors can take many forms, from reluctance to use a health delivery system dominated by English-speaking Whites to preference for alternative healers or clergy more attuned to the culture of the individual. We recommend that future research examine perceptions of barriers as well as use of alternative healers. A more balanced and complete picture would also include attention to protective factors and coping resources related to ethnic identity[31] rather than an exclusive focus on individual psychopathology.

Aspects of the service delivery system play a role as well. The absence of ethnic mental health practitioners is one of the most conspicuous barriers. Also, the intersection of lower social class status and ethnicity, combined with residential segregation, leads to inequities in service delivery that disproportionately affect Blacks and Hispanics. An inner-city resident who has insurance coverage will have more difficulty finding a clinic or mental health practitioner than will a resident of an affluent neighborhood. Finally, an ignorance of the more subtle aspects of cultural and social class differences may contribute to an appearance of insensitivity and thus inhibit the pursuit of care. Qualitative studies are needed to examine the process of mutual estrangement that can result from cultural misunderstandings.

As implied by this study, increasing equity and access under national health insurance will not necessarily close the gap in ethnic use of mental health services. Whether this gap is due to lower availability of services, lower acceptability of services, or both, points to an intriguing point of departure for future research and for planning mental health policies that are effective as well as humane. □

Acknowledgments

This research was supported by National Institute of Mental Health grant MH-46005.

References

1. *Task Panel Reports: The Nature and Scope of the Problem.* Washington, DC: President's Commission on Mental Health; 1978.
2. Shapiro S, Skinner EA, Kessler LG, et al. Utilization of health and mental services: three epidemiologic catchment area sites. *Arch Gen Psychiatry.* 1984;41:971–978.
3. Shapiro S, Skinner EA, Kramer M, Steinwachs DM, Regier DA. Measuring need for mental health services in a general population. *Med Care.* 1985;23:1033–1043.
4. Robins LN, Regier DA, eds. *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study.* New York, NY: Free Press; 1991.
5. Leaf PJ, Bruce ML, Tischler GL, Freeman DH, Weissman MM, Myers JK. Factors affecting the utilization of specialty and general medical mental health services. *Med Care.* 1988;26:9–26.
6. Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK. The de facto US mental and addictive disorders service system. *Arch Gen Psychiatry.* 1993;50:85–94.
7. Snowden LR. Ethnicity and utilization of mental health services: an overview of current findings. In: *Proceedings: Oklahoma Mental Health Research Institute, 1988 Professional Symposium.* Oklahoma City, Okla: Oklahoma Mental Health Research Institute; 1988:227–238.
8. Neighbors HW, Bashshur R, Price R, Selig S, Donabedian A, Shannon G. Ethnic minority mental health service delivery: a review of the literature. *Res Community Ment Health.* 1992;7:53–69.
9. Hough RL, Landsverk JA, Karno M, Burnam A, Timbers DM, Escobar JI. Utilization of health and mental health services by Los Angeles Mexican Americans and non-Hispanic Whites. *Arch Gen Psychiatry.* 1978;44:702–709.
10. Wells KB, Golding JM, Hough RL, Burnam A, Karno M. Factors affecting the probability of use of general and medical health and social/community services for Mexican Americans and non-Hispanic Whites. *Med Care.* 1988;26:441–452.
11. Sussman LK, Robins LN, Earls F. Treatment-seeking for depression by Black and White Americans. *Soc Sci Med.* 1987;24:187–196.
12. Sue S. Community mental health services to minority groups. *Am Psychol.* 1977;32:616–624.
13. Yamamoto J, Silva JA. Do Hispanics underutilize mental health services? In: Gaviria M, Arana JD, eds. *Health and Behavior: Research Agenda for Hispanics.* Chicago, Ill: Simon Bolivar Hispanic American Psychiatric Research and Training Center; 1987.
14. Broman CL. Race differences in professional help seeking. *Am J Community Psychol.* 1987;15:473–480.
15. Scheffler RM, Miller AG. Demand analysis of mental health service use among ethnic subpopulations. *Inquiry.* 1989;26:202–215.
16. Hu T, Snowden LR, Jerrell JM, Nguyen TD. Ethnic populations in public mental health: services choice and level of use. *Am J Public Health.* 1991;81:1429–1434.
17. Lopez S. Mexican-American usage of mental health facilities: under-utilization considered. In: Baron A, ed. *Explorations in Chicano Psychology.* New York, NY: Praeger; 1981:139.
18. Weil TP. Mental health services under a US national health insurance plan. *Hosp Community Psychiatry.* 1991;42:695–700.
19. Andersen R, Newman JF. Societal and individual determinants of medical care utilization in the United States. *Milbank Memorial Fund Q: Health Soc.* 1973;51:95–124.
20. Tweed DL, Goldsmith HF, Jackson DJ, Stiles D, Rae DS, Kramer M. Racial congruity as a contextual correlate of mental disorder. *Am J Orthopsychiatry.* 1990;60:392–400.
21. Knesper DJ, Wheeler JR, Pagnucco DJ. Mental health services: provider's distribution across counties in the US. *Am Psychol.* 1984;39:1424–1434.
22. Wells K, Manning W, Duan J, Ware J, Newhouse J. *Cost Sharing and the Demand for Ambulatory Mental Health Services.* Santa Monica, Calif: Rand Corporation; 1982.
23. Horgan CM. The demand for ambulatory mental health services from specialty providers. *Health Serv Res.* 1986;21:291–298.
24. Taube CA, Kessler LG, Burns BJ. Estimating the probability and level of ambulatory mental health services use. *Health Serv Res.* 1986;21:321–327.
25. Cohen J, Cohen P. *Applied Multiple Regression/Correlation for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1983.
26. Temkin-Greener H, Clark KT. Ethnicity, gender, and utilization of mental health services in a Medicaid population. *Soc Sci Med.* 1988;26:989–996.
27. Vernon SW, Roberts RE. Prevalence of treated and untreated psychiatric disorders in three ethnic groups. *Soc Sci Med.* 1982;16:1575–1582.
28. Woolhandler S, Himmelstein DU. A national health program: northern light at the end of the tunnel. *JAMA.* 1989;262:2136–2140.
29. Bestman E. Blacks and mental health services. In: Silverman WH, ed. *Community Mental Health.* New York, NY: Praeger; 1981.
30. Rogler LH, Malgady RG, Rodriguez O. *Hispanic Mental Health: A Framework for Research.* Malabar, Fla: Krieger Publishing Co; 1989.
31. Vega WA, Rumbaut RG. Ethnic minorities and health. *Ann Rev Sociol.* 1991;17:351–383.

Mental Health Status as a Predictor of Morbidity and Mortality: A 15-Year Follow-Up of Members of a Health Maintenance Organization

Thomas Vogt, MD, MPH, Clyde Pope, PhD, John Mullooly, PhD, and Jack Hollis, PhD

ABSTRACT

Objectives. This study sought to relate preexisting depression, worries, affect balance, and mental health symptoms to subsequent risk of morbidity and mortality.

Methods. A random sample of members of a health maintenance organization were interviewed at home in a 1970/71 survey. Baseline psychological measures from that survey were then related to 15-year mortality and first incidence of cancer, heart disease, hypertension, stroke, functional gastrointestinal disease, and hyperimmune diseases.

Results. Baseline depression and the Langner Mental Health Index predicted incidence of functional gastrointestinal and hyperimmune diseases. The relationship of the Langner index to hyperimmune diseases was particularly strong; mortality, cancer, heart disease, hypertension, and stroke incidence were not related to that index. Except for affect balance, which was worst in the elderly, indications of psychological dysfunction were strongest in the young.

Conclusions. Psychological indices detected increased risk for functional gastrointestinal and hyperimmune diseases but were not related to mortality risk. Further research is needed to disaggregate relationships of the specific conditions that were included in the hyperimmune group. Functional gastrointestinal disease might be preventable with early attention to depressed persons and to those scoring high on the Langner index. (*Am J Public Health.* 1994;84: 227–231)

Introduction

An extensive literature examines the relationship of psychological distress to mortality risk.[1-11] In presenting their own negative data, Roberts et al.[4] surmised that reported associations of psychological state to mortality are misleading because case–control and cross-sectional studies have generally involved persons who are psychologically impaired and thus are not representative of the broader population. Community studies, on the other hand, have been less likely to make such associations.[4-14] Animal studies demonstrate clear relationships between various stresses and immune response, both in vitro and in vivo[15,16]; such effects, if they persist over long periods, might be expected to influence susceptibility to selected morbidities. A meta-analysis of the role of personality in disease causation[17] suggests the possible existence of a "disease-prone" personality that is characterized by anger/hostility and anxiety.

This paper examines the extent to which four baseline measures of mental and emotional health and well-being were able to predict 15-year morbidity and mortality in a random sample of adult enrollees in a large health maintenance organization (HMO). The four measures examined were the Langner Mental Health Index, a depression scale, the Bradburn-Caplowitz Affect-Balance Scale, and the Bradburn Worries Index. The relation of morbidity and mortality in this group to social networks has been examined separately.[18]

Methods

Study Population

The study population was a random sample of 1659 household units with 2573 adult members of the Northwest Region of Kaiser Permanente, a large managed-care organization in Portland, Ore, whose membership is generally representative of the Portland, Ore/Vancouver, Wash, standard statistical metropolitan area.[19] Subjects were respondents to a 1970/71 personal interview survey. Of the households randomly selected, 92% (1529) were successfully interviewed. Thirty individuals who left the health plan less than 6 months after the survey interview were eliminated from the analyses. To be eligible, individuals had to have been enrolled in the health plan prior to January 1, 1969, thus permitting 2 years of retrospective medical records review prior to the survey. Medical records for this population were routinely abstracted from the date of enrollment or January 1967, whichever was earlier.

Independent Variables

The baseline measures were derived from the 1970/71 personal interview survey. These measures include self-reported general health status and scores on the Langner Mental Health Index, a depression scale, the Bradburn-Caplowitz Affect-Balance Scale, and the Bradburn Worries Index.

The Langner index[20] is a count of the number of 22 symptoms (e.g., headaches, night sweats, heart pounding, trouble sleeping, etc.) the respondent often experiences. It is scored from 0 to 22 (mean = 2.19, SD = 2.38). Although de-

The authors are with the Kaiser Permanente Center for Health Research, Portland, Ore.

Requests for reprints should be sent to Thomas M. Vogt, MD, MPH, Kaiser Permanente Center for Health Research, 3800 N Kaiser Center Dr, Portland, OR 97227.

This paper was accepted June 23, 1993.

TABLE 1—Baseline Characteristics of the Study Population (n = 2573)

	No.	%
Age, y		
18–29	398	15.5
30–44	776	30.2
45–59	741	28.8
60–74	529	20.6
75+	129	5.0
Sex		
Male	1187	46.1
Female	1386	53.9
Smoking		
Current	870	33.8
Former	643	25.0
Never	1045	40.6
Missing	15	0.6
Socioeconomic status (self-reported)		
Upper	26	1.0
Upper middle	357	13.9
Middle	1385	53.8
Working	721	28.0
Lower	22	0.9
Missing	62	2.4
Subjective health status		
Excellent	888	34.5
Good	1171	45.5
Fair/poor	461	17.9
Missing	53	2.1

Note. Percentages may not add to 100 because of rounding.

signed as a general measure of mental health, it is widely regarded as an index of psychophysiological symptomatology and general malaise.[21]

The depression index was created from questions on the survey that allowed a close approximation of the clinical definition of depression in the American Psychiatric Association's *Diagnostic and Statistical Manual*, 3rd edition.[22] The instrument and distributions of scores on its component questions are discussed in McFarland et al.[23]

The Bradburn-Caplowitz Affect-Balance Scale[24] was created from questions asking about five negative and five positive emotional states (e.g., during the past few weeks, did you ever feel proud? lonely?). To an initial score of 5, +1 is added for each "yes" response to a positive emotional state and −1 for each "yes" response to a negative emotional state. Scores range from 0 to 10, a higher score indicating a positive affect (mean = 7.22, SD = 1.79).

The Bradburn Worries Scale[25] consists of 12 items (e.g., during the past few weeks, have you been worried about money? job? world?) weighted from not worried (0) to very worried (2). Scores range from 0 to 24 (mean = 3.48, SD = 2.66). This index is intended to reflect transitory mental states (unlike the Langner index, which purportedly measures more enduring states).

Dependent Variables

Vital status as of the end of 1985 was determined for all members of the study population through searches of the national death index and vital records for Oregon and Washington states.

International Classification of Diseases, version 7 (ICD-7)[26] codes were used to classify all confirmed morbidities for all years of the study. The use of a single coding system prevented discontinuities secondary to ICD code revisions.

The first incidence was determined for the following morbid outcome groups (ICD-7 codes): ischemic heart disease (420, 421.5, 422.1), cancer (140–205, excluding 191), stroke (330–334, 335.1), hypertension (447.0, 447.2), infection (002–138), diseases of hyperimmunity (allergy [240, 242–245], asthma [241], osteoarthritis [723], autoimmune diseases [252.0, 254.0, 256, 271.1, 290.0, 296, 456, 704, 744]), and functional gastrointestinal diseases (540–544, 571–573). Persons with a past history of a morbid condition were excluded from analyses of that condition so that the outcome was always defined as the first incidence of a morbidity.

Analyses

Cox proportional hazards regression models were used to assess the 15-year relative hazards of mortality and morbidity. These nonparametric models allow the hazards of death and disease incidence to include follow-up time but require that the proportional effects of predictors on hazards be constant over the entire follow-up interval. The Cox regression procedure allows censoring due to health plan disenrollment and includes a goodness-of-fit test of the proportional hazards assumption. The exponential function of a Cox regression coefficient is the relative hazard of the outcome for a unit difference in the predictor variable.

Analyses were adjusted for age, sex, self-reported health status, self-reported social class, cigarette smoking status, and duration of health plan membership. For most of the analyses, the predictor variables were divided into approximate tertiles, with the upper (high) tertile serving as the referent for comparisons with the bottom (low) tertile. The SAS SUGI procedure PHGLM[27] was used.

Kaplan Meier and log-log hazard plots and proportional hazard z scores were used to verify Cox model assumptions. Stepwise inclusion criteria were used to identify interactions of age, gender, and baseline subjective health status that significantly improved the predictive power of the main effects models.

Results

Table 1 shows the characteristics of the study population. The population is a broadly distributed group of HMO members generally representative of the metropolitan area from which they were drawn.

Table 2 presents the percentage of individuals within each of five age groups that fell into the approximate tertiles of the various psychosocial measures. For each measure, cut points were selected that most nearly distributed the respondents into thirds. Depression scores were highest in those under 30, followed by those over 75. Those from 30 to 74 years of age were least likely to be depressed. Persons under 30 were least likely to fall in the lowest (most healthy) tertile on the Langner scale and most likely to fall in the highest tertile. Persons aged 60 to 74 were most likely to fall in the lowest tertile. The worries index was inversely related to age, with the two oldest age groups (60 and above) showing the fewest worries. Persons over age 75 were most likely to fall in the lowest tertile of affect balance (worst affect balance), and those aged 30 to 44 had the highest affect-balance.

Table 3 presents the Pearson correlations among the four psychological scales. Depression and the Langner index were highly correlated (r = .83). The other measures were significantly but more weakly correlated.

Table 4 presents the 15-year relative hazards (RHs) of mortality and morbidity incidence. For each index, the lowest tertile was compared with the highest tertile. The table includes both unadjusted rates and rates adjusted for age, sex, smoking status, socioeconomic status (SES), self-reported health status, and duration of health plan membership.

Those in the low and medium tertiles of the depression index at baseline (not depressed) were at increased risk for subsequent hypertension compared with those scoring high on the depression index. Depression was also related to an increased incidence of functional gastrointestinal disease and was marginally as-

sociated with an increased hazard for diseases of hyperimmunity. The other five dependent variables were not related to depression. For men only (data not shown), the risk of dying during the follow-up interval was lower for the least depressed compared with the most depressed tertiles (RH = 0.71; confidence interval [CI] = 0.52, 0.97). No other interactions with sex were observed.

A high score on the Langner index was related to an increased risk of both functional gastrointestinal and hyperimmune diseases among all persons. The index interacted with age in predicting incidence of hypertension. Younger individuals with a low mental health score were at increased risk for hypertension (RH = 1.47; CI = 1.05, 2.04 at age 30), while older persons with a low mental health score were at reduced hazard (RH = 0.76; CI = 0.62, 0.92). No other main effects or interactions were noted.

A high score on the worries index was associated with an increased risk for functional gastrointestinal disease. The worries index was also strongly related to mortality and to incidence of ischemic heart disease, cancer, hypertension, and stroke in the unadjusted analyses. Adjustment for age, sex, smoking, health status, SES, and duration of health plan membership abolished these relationships.

The unadjusted affect balance score was significantly associated with mortality, ischemic heart disease, stroke, and functional gastrointestinal disease. After adjustment, the affect balance scale was not significantly associated with any of these outcomes, but low affect balance was marginally related to elevated risks of death and stroke.

To assess the degree to which the addition of the mental health variables improved the fit of a model involving only covariates, we reran all the models without the mental health variables (i.e., with covariates only). We then used the likelihood ratio test to determine whether the main effects models of covariates alone were significantly improved by the addition of the four mental health indices. Addition of the depression, Langner mental health, and worries indices to the model produced highly significant improvements in the model for all outcomes except in the case of the worries index as a predictor of stroke. *P* values associated with these chi-squares were extremely small—in most instances, much less than .0001. The affect balance scale was not associated with a significant improvement in the model for any of the outcomes except cancer incidence ($P < .005$).

TABLE 2—Percentage of Participants, Distributed by Age, within Tertiles of Psychological Scales

Index/Tertile	Age, y <30	30–44	45–59	60–74	75+
Depression					
1 (least)	20.8	34.9	36.1	35.0	24.4
2	27.1	30.9	31.2	29.2	32.2
3 (most)	52.2	34.2	32.7	35.8	43.3
Langner Mental Health Index					
1 (best)	21.5	28.3	25.3	30.7	25.3
2	39.2	40.0	43.5	35.3	39.0
3 (worst)	39.2	31.7	31.2	34.0	35.8
Bradburn Worries Scale					
1 (fewest)	21.2	32.1	44.6	56.7	59.1
2	26.9	27.1	30.2	27.2	24.8
3 (most)	51.9	40.8	25.2	16.1	16.2
Bradburn-Caplowitz Affect-Balance Scale					
1 (worst)	32.7	27.1	30.0	36.1	41.8
2	41.2	43.2	42.7	42.9	40.9
3 (best)	26.1	29.5	27.4	21.0	17.3

TABLE 3—Pearson Intercorrelations of Psychological Measures*

	Langner Mental Health Index	Bradburn Worries Scale	Bradburn-Caplowitz Affect-Balance Scale
Depression	.83	.42	−.55
Langner Mental Health Index		.37	−.45
Bradburn Worries Scale			−.21

*$P < .001$.

Discussion

These analyses show a strong prospective association between the mental health, depression, and the worries indices, on the one hand, and the first incidence of functional gastrointestinal disease and a group of diseases we classified as diseases of hyperimmunity, on the other. Unfortunately, this latter grouping was created for other reasons, and we were not able to break up that group into its major component parts (allergies/hay fever, asthma, osteoarthritis, autoimmune diseases). We plan to do so in the future. In the adjusted models, psychological indices were not related to mortality hazard (except possibly for depression among men only) or to the incidence of major morbidities such as ischemic heart disease and cancer. The unadjusted worries and affect balance indices were strongly related to these outcomes, however, emphasizing the importance of appropriate adjustments when analyzing the relationships of psychological variables to morbidity and mortality.

The relationship of three of the four indices to functional gastrointestinal and hyperimmune conditions is intriguing. Functional gastrointestinal diseases obviously have a strong psychological component. But, to our knowledge, this is the first evidence that a depression index, the Langner Mental Health Index, and the Bradburn Worries Scale can be used to identify those at high risk for future functional gastrointestinal disease. A 1957 study of army recruits[28] probably gave rise to the general perception that personality and ulcer are linked, but no well-controlled studies exist that confirm this commonly held perception. Similarly, arthritis, particularly rheumatoid arthritis, is often perceived to be related to psychological factors. Persons with rheumatoid arthritis have more psychological problems than controls and appear to have more psychological stress.[29] Certainly, rheumatoid arthritis is involved with immune response. But prospective studies

TABLE 4—Fifteen-Year Relative Hazard (RH) of Mortality and Morbidity, by Mental Health, Depression, Worries, and Affect-Balance Scores[a]

Outcome	Unadjusted			Adjusted		
	RH	95% CI	P	RH	95% CI	P
Death						
Mental health	1.05	0.83, 1.34	.67	1.03	0.79, 1.34	.85
Depression	0.96	0.76, 1.20	.71	0.91	0.71, 1.17	.47
Worries	2.19	1.73, 2.78	<.001	0.94	0.74, 1.21	.64
Affect balance	1.99	1.54, 2.58	<.001	1.25	0.95, 1.64	.10
Ischemic heart disease						
Mental health	1.04	0.76, 1.42	.67	0.88	0.62, 1.24	.46
Depression	1.13	0.86, 1.49	.39	0.94	0.70, 1.28	.71
Worries	2.34	1.72, 3.20	<.001	1.18	0.85, 1.63	.32
Affect balance	1.46	1.08, 1.98	.01	1.13	0.82, 1.54	.47
Cancer						
Mental health	1.00	0.70, 1.43	1.00	1.01	0.69, 1.49	.94
Depression	1.08	0.79, 1.49	.60	1.08	0.77, 1.52	.66
Worries	1.63	1.17, 2.27	<.001	0.99	0.70, 1.40	.95
Affect balance	1.17	0.83, 1.65	.36	1.00	0.70, 1.43	1.00
Hypertension						
Mental health	1.13	0.90, 1.43	.29	1.08	0.84, 1.39	.56
Depression	1.43	1.15, 1.77	<.001	1.37	1.09, 1.73	<.001
Worries	1.42	1.14, 1.75	<.001	1.17	0.93, 1.46	.18
Affect balance	1.01	0.80, 1.27	.94	0.99	0.78, 1.26	.94
Stroke						
Mental health	0.83	0.57, 1.21	.32	0.83	0.55, 1.25	.37
Depression	0.85	0.60, 1.20	.35	0.84	0.57, 1.22	.34
Worries	1.98	1.38, 2.84	<.001	0.91	0.63, 1.33	.64
Affect balance	2.09	1.41, 3.10	<.001	1.44	0.95, 2.19	.08
Functional gastrointestinal diseases						
Mental health	0.62	0.49, 0.79	<.001	0.71	0.55, 0.92	.01
Depression	0.63	0.51, 0.79	<.001	0.68	0.54, 0.87	<.001
Worries	0.76	0.62, 0.94	.01	0.71	0.57, 0.89	<.001
Affect balance	1.27	1.01, 1.60	.04	1.19	0.94, 1.51	.15
Hyperimmune diseases						
Mental health	0.63	0.51, 0.78	<.001	0.67	0.53, 0.84	<.001
Depression	0.84	0.69, 1.01	.06	0.84	0.68, 1.03	.09
Worries	1.11	0.92, 1.33	.29	0.87	0.71, 1.07	.18
Affect balance	1.07	0.87, 1.31	.54	0.98	0.79, 1.21	.84

[a]Low tertile relative to high tertile; adjusted for age, sex, socioeconomic status, length of health plan membership, subjective health status, and smoking.

on these issues are woefully lacking. Is it possible to intervene with appropriate therapy for those who score high on these indices and thereby to reduce future functional gastrointestinal conditions and/or conditions associated with excessive inflammatory/immune response? This question deserves further attention.

There are well-described biochemical pathways that might link hypothalamic-adrenocortical hormones to the risk of heart disease and cancer.[30,31] Despite the common, yet relatively undocumented, knowledge that periods of stress are related to exacerbations of asthma, allergies, and arthritis, there is no comparable biological model to account for the relationships between psychological factors and excessive immune response. There have been observations that persons with allergies are at decreased risk for cancer,[32] but little has been done to specify whether (or how) psychological states might be involved. The present study suggests that perhaps these morbidities are linked to psychological states. This study has also examined the relationships of social networks[18] and other psychosocial variables to these outcomes.

Although the depression and mental health indices are highly correlated, only the former was inversely related to the risk of newly diagnosed hypertension. Physicians may be less likely to diagnose asymptomatic conditions such as hypertension in depressed patients than in nondepressed patients, and/or depression may be associated with reductions in blood pressure. There is no basis for attributing this inverse relationship to antidepressant medications. Such medication might, in fact, be expected to increase blood pressure.

Although affect balance, a measure of well-being, tended to be negatively related to the risk of death and stroke, it was not significantly associated with any of the study outcomes after adjustment for covariates.

Persons under 30 years of age were more depressed and more worried and had more long-term mental health symptoms than other age groups. Only affect-balance was worse among the old than the young. The high completion rate (92%) of the random sample suggests that these findings are not the result of a healthy volunteer effect and that age does not lead to a general deterioration in these aspects of mental health. Indeed, some aspects appear to improve with age.

Of course, many comparisons were performed in conducting these exploratory analyses, and some significant associations might be expected by chance alone. However, the associations reported here are very powerful, and few are likely to have resulted from chance. It is possible that specific morbid conditions present at baseline may have confounded the observed relationships. Further exploration of the relationships of these mental health indices to morbid conditions would help to detect such problems.

Further exploration of the relationships between psychological states and the future risk of functional gastrointestinal and hyperimmune conditions is warranted. In particular, the relationship of these measures to the subcomponent conditions that make up the hyperimmune group would be of interest. □

Acknowledgments

This work was supported by National Institute on Aging grant R01 AG05682.

The authors thank Denise Ernst, Lisa Sha, Elizabeth O'Connor, and Mary Longacre for their contributions.

References

1. Lieberman AA, Coburn AF. The health of the chronically mentally ill: a review of the literature. *Community Ment Health J.* 1986;22:104–116.
2. Tsuang MT, Simpson JC. Mortality studies in psychiatry: should they stop or proceed? *Arch Gen Psychiatry.* 1985;42:98–103.
3. Zilber N, Schufman N, Lerner Y. Mortality among psychiatric patients—the groups at risk. *Acta Psychiatr Scand.* 1989;79:248–256.
4. Roberts RE, Kaplan GA, Camacho TC. Psychological distress and mortality: evi-

dence from the Alameda County study. *Soc Sci Med.* 1990;31:527–536.
5. Rorsman B, Hagnell O, Lanke J. Violent death and mental disorders in the Lundy study. Accidents and suicides in a total population during a 25-year period. *Neuropsychobiology.* 1982;8:233–240.
6. Rorsman B, Hagnell O, Lanke J. Mortality in the Lunby study: natural death in different forms of mental disorder in a total population investigated during a 25-year period. *Neuropsychobiology.* 1982;8:188–197.
7. Singer E, Garfinkel R, Cohen SM, Srole L. Mortality and mental health: evidence from the Midtown Manhattan restudy. *Soc Sci Med.* 1976;10:517–525.
8. Murphy JM, Monson RR, Olivier DC, Sobol AM, Leighton AH. Affective disorders and mortality: a general population study. *Arch Gen Psychiatry.* 1987;44:473–480.
9. Bruce ML, Leaf PJ. Psychiatric disorders and 15-month mortality in a community sample of older adults. *Am J Public Health.* 1989;79:727–730.
10. Somervell PD, Kaplan BH, Heiss G, Tyroler HA, Kleinbaum DG, Obrist PA. Psychologic distress as a predictor of mortality. *Am J Epidemiol.* 1989;130:1013–1023.
11. Weissman MM, Myers JK, Thompson WK, Belanger A. Depressive symptoms as a risk factor for mortality and for major depression. In: Erlenmeyer-Kimling L, Miller N, eds. *Life-Span Research and the Prediction of Psychopathology.* Potomac, Md: Erlbaus; 1986:251.
12. Hankin JR, Locke BZ. The persistance of depressive symptomatology among prepaid group practice enrollees: an exploratory study. *Am J Public Health.* 1982;72:1000–1007.
13. Hornstra RK, Klassen D. The course of depression. *Compr Psychiatry.* 1977;18:119–125.
14. Lin N, Ensel WM. Depression-mobility and its social etiology: the role of life events and social support. *J Health Soc Behav.* 1984;25:176–188.
15. Ader R, Cohen N. CNS-immune system interaction: conditioning phenomena. *Behav Brain Sci.* 1985;8:379–394.
16. Vogt TM. Stress and illness: psychobiological links. In: Ory MG, Abeles RP, Lipman PD, eds. *Aging, Health and Behavior.* Newbury Park, Calif: Sage Publications; 1992;207–238.
17. Friedman HS, Booth-Kewley S. The "disease-prone personality": a meta-analytic view of the construct. *Am Psychol.* 1987;42:539–555.
18. Vogt TM, Mullooly JP, Ernst D, Pope CR, Hollis JF. Social networks as predictors of incidence, survival, and mortality from ischemic heart disease, cancer, stroke, and hypertension. *J Clin Epidemiol.* 1992;45:659–666.
19. Greenlick MR, Freeborn DK, Pope CR. *Health Care Research in an HMO: Two Decades of Discovery.* Baltimore, Md: The Johns Hopkins University Press; 1988:13–14.
20. Srole L, Langner TS, Michael ST, et al. *Mental Health in the Metropolis: The Midtown Manhattan Study.* New York, NY: McGraw-Hill; 1962; 1.
21. Seiler LH. The 22-item scale used in field studies of mental illness: a question of method, a question of substance, and a question of theory. *J Health Soc Behav.* 1973;14:252–264.
22. *Diagnostic and Statistical Manual of Mental Disorders.* 3rd ed. Washington, DC: American Psychiatric Association; 1980.
23. McFarland BH, Freeborn DK, Mullooly JP, Pope CR. Utilization patterns among long-term enrollees in a prepaid group practice health maintenance organization. *Med Care.* 1985;23:1221–1233.
24. Bradburn NM. *The Structure of Psychological Well-Being.* Chicago, Ill: Aldine Press; 1969.
25. Bradburn NM, Caplovitz D. *Reports on Happiness.* Chicago, Ill: Aldine Press; 1965.
26. US Dept of Health and Human Services, National Center for Health Statistics. *International Classification of Diseases*, version 7. Washington, DC: US Govt Printing Office; 1962.
27. SAS Institute, Inc. *SUGI Supplemental Library User's Guide, Version 5 Edition.* Cary, NC: SAS Institute; 1986.
28. Weiner H, Thaler M, Reiser MG, Mirsky IA. Etiology of duodenal ulcer. *Psychosom Med.* 1957;19:1–10.
29. Anderson KO, Bradley LA, Young LD, McDaniel LK, Wise CM. Rheumatoid arthritis: review of psychological factors related to etiology, effects, and treatment. *Psychol Bull.* 1985;98:358–387.
30. Troxler RG, Sprague EA, Albanese RA, Fuchs R, Thompson AJ. The association of elevated plasma cortisol and early atherosclerosis as demonstrated by coronary angiography. *Atherosclerosis.* 1977;26:151–162.
31. Rogers MP, Dubey D, Reich P. The influence of the psyche and the brain on immunity and disease susceptibility: a critical review. *Psychosom Med.* 1979;41:147–164.
32. Vena JE, Bona JR, Byers TE, Middleton E Jr, Swanson MK, Graham S. Allergy-related diseases and cancer: an inverse association. *Am J Epidemiol.* 1985;122:66–74.

Dementia-Free Life Expectancy in France

Karen Ritchie, PhD, Jean-Marie Robine, PhD, Luc Letenneur, PhD, and Jean-François Dartigues, PhD

ABSTRACT

Objectives. Increasing concern with the quality of gains in life years has led to the development of a new synthetic indicator of population health:health expectancy. Until now, calculations have been made for physical disabilities only. A first estimate of mental health expectancy is presented: dementia-free life expectancy.

Methods. Sullivan's method was used to calculate dementia-free life expectancy for a random representative sample of 4134 persons over 65 years of age in the Bordeaux region of France. The diagnosis of senile dementia was made in two stages, based on *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) criteria.

Results. At 65 years of age, a person's dementia-free life expectancy is 16.9 years within a total life-expectancy of 17.7 years; it decreases with age in parallel with the decrease in total life expectancy so that life expectancy with dementia stays constant at 0.8 years. Although dementia prevalence increases with age, if the prevalence is adjusted for mortality, the largest number of persons with dementia are in their early eighties. At each age women have a higher dementia-free life expectancy.

Conclusions. Trends in dementia-free life expectancy are similar to those found in disability-free life expectancy. Because the dementia prevalence rates used in this estimate resemble a general model derived from meta-analysis, it can be assumed that similar results will be found in other Western countries with similar mortality rates. (*Am J Public Health.* 1994;84:232–236)

Introduction

Perhaps one of the most striking recent changes in the field of public health has been the appreciation that health is more than survival and that health care policy should look beyond extending life to assess the impact of disease both on the individual's capacity to lead a normal life and on the society that cares for the individual. This change in emphasis, in conjunction with the observation that increases in life expectancy appear to be accompanied by increases in the prevalence of chronic disease and disability, has led to widespread interest in the concept and calculation of "healthy life expectancy." Studies of health expectancy examine changes in mortality in relation to the prevalence of physical disease and disability, thus addressing the question of whether the years we have gained as a result of preventive and therapeutic health programs are years of good or poor health.

A large number of health expectancy calculations have now been made in both developed and developing countries[1,2]; however, these have all focused on the measurement of physical disability. Mental disorder has received little attention, although a strong motivating force for calculations of health expectancy has been the theory of Kramer, who predicted in 1980 a "coming pandemic of mental disorders and associated chronic diseases and disabilities."[3] Kramer's work, which is principally based on the results of population studies of senile dementia and schizophrenia, clearly targets mental disorder as the primary area of concern. Although the entire range of mental disorders has yet to be explored in the context of health expectancy calculations, life expectancy free of senile dementia (referred to here as dementia-free life expectancy) is perhaps the most urgent, as the prevalence of dementia increases with age[4] and considerable concern has been expressed by health professionals that increased life expectancy, particularly among the oldest of the old, will lead to an epidemic of senile dementia by the end of the century.

In France as in other Western countries, approximately 50% of long-stay hospital beds are presently occupied by elderly persons with senile dementia. This condition involves progressive and irreversible deterioration of all cognitive functions, with rapid loss of autonomy and independence. The burden for both professional and family caregivers is considerable: the average length of survival from dementia onset is 8 years,[5] no specific medical treatment is presently available, and persons with dementia have increased susceptibility to a wide range of physical disorders.[6] Progressive loss of language accompanied by failure to recognize family and friends lead to rapid social exclusion. If the aim of health expectancy calculations is to determine the quality of gains in life expectancy, then age-related intellectual deterioration cannot be overlooked; it will affect not only the elderly person but also the person's caregivers, who themselves will be at increased risk of both physical and mental ill health.[7]

Karen Ritchie and Jean-Marie Robine are with the French National Institute of Health and Medical Research (INSERM), Montpellier, France. Luc Letenneur and Jean-François Dartigues are with the Université de Bordeaux, INSERM, Bordeaux, France.

Requests for reprints should be sent to Karen Ritchie, PhD, INSERM Equipe "Vieillissement Cognitif," Centre Val d'Aurelle, Parc Euromédecine, 34094 Montpellier Cedex 5 France.

This paper was accepted June 23, 1993.

Given that functional loss is one of the principal consequences of dementia, it is reasonable to question the value of measuring dementia-free life expectancy independently of life expectancy without physical disability. There are several reasons for making such a distinction. First, there is the extent of the loss; senile dementia involves a rapid and global loss of function, and the services provided to physically disabled persons are often inappropriate. Elderly persons with senile dementia not only have difficulty in carrying out tasks but may attempt to perform tasks at inappropriate times. Admissions to long-term care thus tend to be related less to degree of functional loss than to problems of supervision.[7,8] In addition, persons with senile dementia are unable to appreciate the implications of their disorder or make decisions concerning their own care. Health services must therefore assume not only the physical care but also the decision-making functions of the person.

It is the aim of the present paper to provide a first estimation of life expectancy without irreversible progressive intellectual deterioration, or dementia-free life expectancy. The calculation is made for the French population in 1990. The data and methods of calculation used are described in sufficient detail to permit researchers in other countries to replicate the calculations where similar data is available.

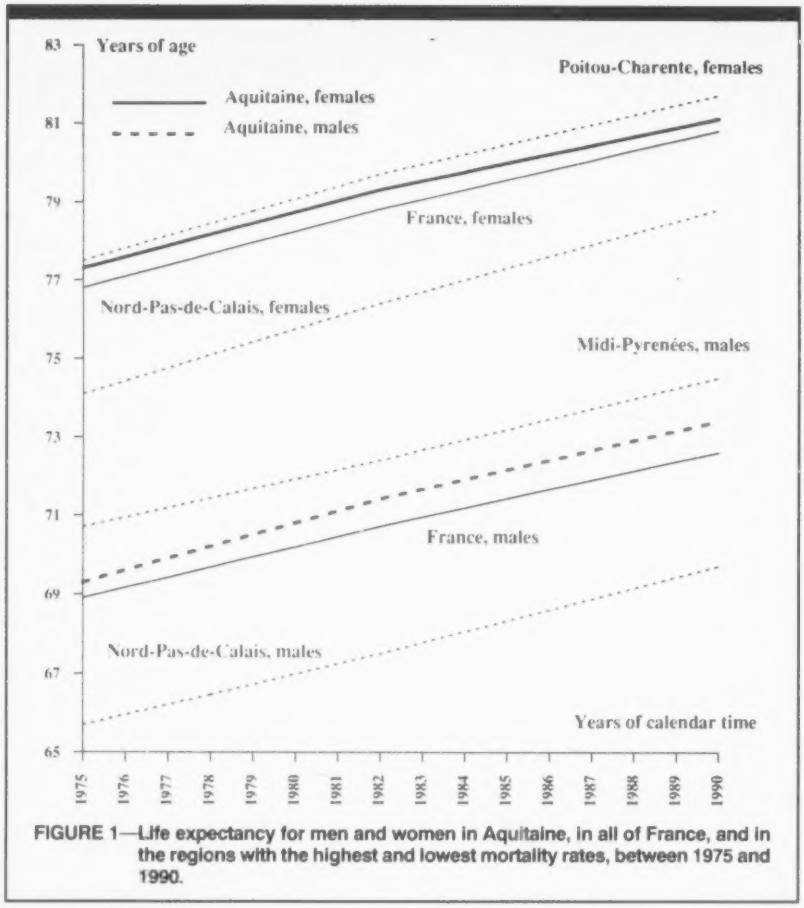


FIGURE 1—Life expectancy for men and women in Aquitaine, in all of France, and in the regions with the highest and lowest mortality rates, between 1975 and 1990.

Methods

Data Sources

Senile dementia prevalence data. Undoubtedly, the lack of interest in dementia-free life expectancy studies to date has been due largely to the difficulties inherent in the detection of senile dementia in the general population. These problems have been discussed in greater detail in relation to health expectancy studies in a previous communication.[9] The development of internationally recognized diagnostic criteria for senile dementia such as DSM-III-R[10] and ICD-9[11] have permitted epidemiologists since the 1980s to report relatively consistent prevalence rates for moderate to severe dementia of around 5% over age 65 across national boundaries.

There has, however, been considerable variation in the reported rates of mild dementia, which have ranged from 2.6% in Britain and Scandanavia[12] to 52.7% in Japan,[13] suggesting continued difficulty in the early recognition of the disorder. Epidemiological studies therefore tend at present to focus on moderate to severe dementia as the resultant prevalence estimates are more accurate and consistent than those for mild dementia. In a meta-analysis[4] of 13 general population studies conducted in Europe, the United States, Canada, and Asia, and employing internationally recognized diagnostic criteria of senile dementia, we were able to derive the following general logistic model of the relationship between age and prevalence:

$$\log_e(p) = -12.43(0.62) + 0.121(0.00785)\ \text{Age},$$

where $\log_e(p)$ represents the proportion of case. No significant differences were found among the 13 studies for either slopes or intercepts; studies using DSM-III-R criteria were found to provide the best fit.

The senile dementia prevalence data used for the present dementia-free life expectancy calculation come from a study of normal and pathological brain aging conducted in the Bordeaux region of France.[14] The project, known as PAQUID (Personnes Agées Quid), is a prospective study of a representative random sample from the electoral roli of 3777 persons older than 65 years living independently in two administrative departments of Aquitaine. The PAQUID study is supplemented by a random sample of 357 elderly persons living in institutions (the PAVIN [Personnes Agées Vivant en Institution] program). The institutionalized population is overrepresented, as the percentage of elderly persons living in institutions in the area is only 6%. A weighting system is therefore used in the calculation of prevalence rates to adjust for this overrepresentation. The samples are stratified by age, sex, and size of the community.

The diagnosis of dementia is made in two stages: systematic neuropsychometric screening of all subjects by a psychologist using DSM-III-R criteria, followed by confirmation of positive cases by a neurologist. The prevalence rates derived from this study show a good fit with the model derived from the meta-analysis, suggesting that the PAQUID–PAVIN data are consistent with rates reported

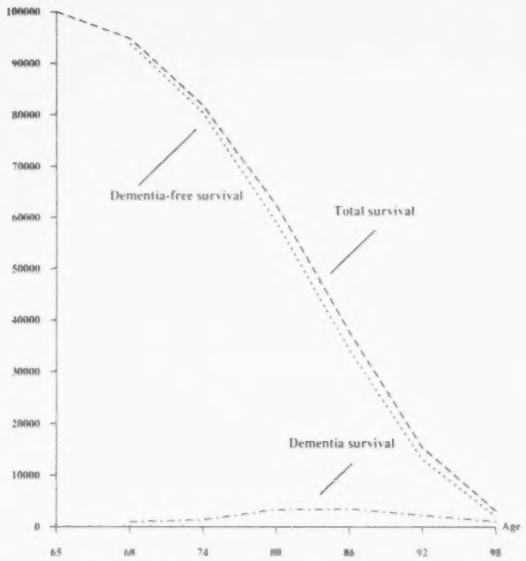


FIGURE 2—Proportion of dementia-free survivors and those with dementia, between the ages of 65 and 98 years, estimated for metropolitan France, 1988–1990 (both sexes combined).

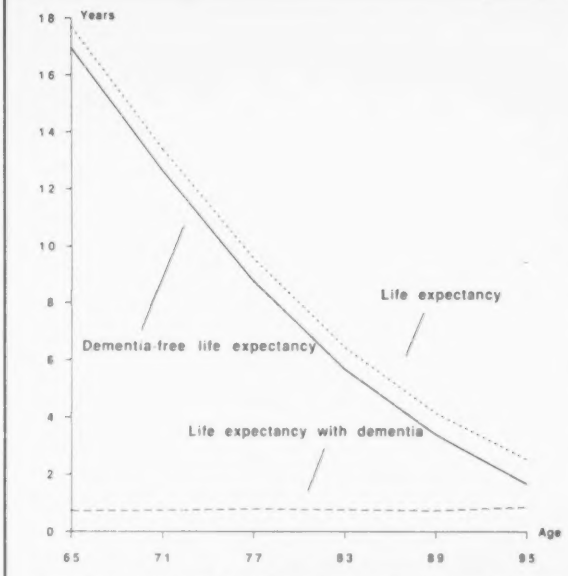


FIGURE 3—Dementia-free life expectancy for persons between the ages of 65 and 95 years, as calculated by Sullivan's method in comparison with total life expectancy and life expectancy with dementia, France, 1988–1990 (both sexes combined).

elsewhere. Although PAQUID and PAVIN are longitudinal studies, a complete data set is presently available only for the first year. These cross-sectional data were collected between January 1988 and October 1990. The prevalence estimates are for moderate to severe dementia; possible mild dementia has been excluded.

The observed percentage prevalence of senile dementia in Aquitaine is derived from the weighted combination of observations from PAQUID and PAVIN. The data were broken down by 6-year age groups because our previous meta-analysis indicated that senile dementia rates double every 6 years. The following prevalence rates were found (number of persons in each age group are given in parentheses for PAQUID/PAVIN): 1.06% between ages 65 and 70 years (953/35), 1.67% between ages 71 and 76 years (730/45), 5.41% between ages 77 and 82 years (659/77), 9.33% between ages 83 and 88 years (330/112), 15.01% between ages 89 and 94 years (97/71), and 34.13% between ages 95 and 100 years (11/16).

Mortality data. A life table for France for the years 1988 through 1990 for each year of age between 65 and 100 years for both sexes combined was provided by the National Institute for Statistics and Economic Studies.[15] After each census the Institute also publishes mortality data for each of the regions of France. Provisional life expectancies at birth are presently available for the years 1989 and 1990 combined[16,17]; from these data we were able to ascertain that life expectancy in the Aquitaine region is one of the highest in France, for both men and women. This is seen in Figure 1, which situates Aquitaine in relation to France as a whole and in relation to the regions with the lowest life expectancy (Nord-Pas-de-Calais) and the highest life expectancy (Poitou for women and Midi-Pyrenées for men).

If mortality rates remain stable, Aquitaine's current rates reflect the situation for the whole of France in about 3 years. For the purposes of this study, the Institute has also provided us with provisional mortality quotients for the Aquitaine region at 1-year intervals from 65 to 97 for both sexes combined calculated from the 1989 and 1990 data. Both national and regional mortality data are therefore combined with the regional senile dementia prevalence data. While it is desirable to combine mortality and morbidity data from the same region, the larger numbers obtained from the national estimates are likely to be less subject to sampling bias. We are also primarily interested in the calculation of national estimates. Mortality rates are summarized in a single figure, that of life expectancy at age 65 years (or mean duration of life after age 65 years). This is the mean of the ages at death of a fictitious generation of persons older than 65 years within the mortality conditions in France and Aquitaine.

Calculation Methods

Sullivan's method[18] has been chosen for the estimation of dementia-free life expectancy. It is the only method that permits the use of prevalence data in the calculation of health expectancies, a characteristic that accounts for the method's widespread use.[1] As a result of its popularity it is probably the method whose limits and biases are best understood. The principle behind Sullivan's method is very straightforward. From a period life table a calculation is made of the number of years lived between different ages by a fictitious cohort living all its life in the conditions of the period. Then the observed prevalence rates at each age or by age group are used to estimate the number of years lived with or without the specific morbid condition.

Results

For persons in metropolitan France aged 65 years between 1988 and 1990, life expectancy for both sexes combined is 17.7 years, giving a mean age at death of 82.7 years. The proportion of nondemented survivors after age 65 also decreases regularly in parallel with the number of survivors (see Figure 2). The

proportion of survivors is 75% at 76 years, 50% at 83 years, and 25% at 88 years; the proportion of nondemented survivors is estimated at 75% at 75 years, 50% at 82 years, and 25% at 87 years. There is thus a gap of just under a year for the entire length of the survival curve. The proportion of demented survivors in relation to the total number in the initial cohort increases from 1% at 68 years to reach a plateau of 3.5% at 80 through 86 years, then decreases steadily to 2.5% at 92 years and almost 1% at 98 years. This observation, illustrated in Figure 2, is due to a decrease with age due to mortality concurrent with the increase with age of the number of persons becoming demented. The maximum number of survivors in the population with dementia are found between the ages of 80 and 86 years (Figure 2); that is, the modal or largest number of persons with dementia are in their mid-80s. We might conclude from this that although the prevalence of dementia increases with age, when we adjust for mortality we see that the greatest burden in terms of health care provision will be for persons in their early 80s.

At 65 years dementia-free life expectancy is 16.9 years, within a total life expectancy of 17.7 years; at 77 years, dementia-free life expectancy is 8.7 years within a total life expectancy of 9.6 years; and at 89 years, it is 3.4 years within a total life expectancy of 4.2 years. As for the proportion of survivors, dementia-free life expectancy decreases regularly with age in parallel with the decrease in total life expectancy, with the rather curious result (see Figure 3) that life expectancy with dementia (not to be confused with average duration of survival of persons who become demented) does not vary with age after 65 years but stays constant at around 0.8 years. Between 65 and 95 years the survivors at each age have a continual expectation of living 0.8 years with dementia.

Life expectancy without dementia takes up 96% of life expectancy at 65 years, 91% at 77 years, and 81% at 89 years. At 95 years, life expectancy without dementia still represents only 67% of life expectancy, a figure that is probably artificially low owing to small numbers. This suggests that the proportion of life years spent in dementia grows with the number of years lived; however, the calculation of the regression line for each 6-year age group indicates that we would have to go beyond 170 years of age before dementia represented the totality of residual life expectancy. In other words, contrary to the fears commonly expressed by geriatricians, it is highly unlikely that all persons will be demented even at a very advanced age.

TABLE 1—Dementia-Free Life Expectancy and Dementia-Free Life Expectancy as a Proportion of Life Expectancy[a] from age 65, France, 1988–1990, and the Aquitaine Region, 1989–1990

Age, y	Dementia-Free Life Expectancy, y	Dementia Life Expectancy, y	Total Life Expectancy, y	Proportion of Life Expectancy Dementia-Free, %
		Men, France		
65	14.8	0.6	15.4	96.4
71	11.0	0.6	11.5	95.2
77	7.6	0.6	8.2	92.6
83	5.0	0.6	5.6	89.4
89	3.1	0.6	3.7	84.5
95	2.3	0.0	2.3	100.0
		Men, Aquitaine		
65	15.2	0.6	15.8	96.3
71	11.2	0.6	11.8	95.2
77	7.7	0.6	8.3	92.5
83	4.9	0.6	5.5	89.3
89	3.0	0.6	3.6	84.2
95	2.1	0.0	2.1	100.0
		Women, France		
65	18.8	0.9	19.7	95.4
71	13.9	0.9	14.9	93.7
77	9.5	1.0	10.5	90.8
83	6.0	0.9	6.9	87.1
89	3.5	0.9	4.4	80.4
95	1.6	1.0	2.6	60.8
		Women, Aquitaine		
65	19.1	0.9	20.1	95.3
71	14.2	1.0	15.1	93.6
77	9.7	1.0	10.7	90.9
83	6.1	0.9	7.0	87.2
89	3.5	0.8	4.3	80.8
95	1.4	0.9	2.3	60.8
		Overall, France		
65	16.9	0.8	17.7	95.8
71	12.6	0.8	13.4	94.2
77	8.7	0.8	9.6	91.4
83	5.6	0.8	6.4	87.7
89	3.4	0.8	4.2	81.2
95	1.7	0.8	2.5	66.8
		Overall, Aquitaine		
65	17.3	0.8	18.1	95.7
71	12.9	0.8	13.7	94.2
77	8.9	0.8	9.7	91.4
83	5.7	0.8	6.5	87.8
89	3.4	0.8	4.1	81.4
95	1.5	0.8	2.3	66.8

[a]As estimated by Sullivan's method.

Values of dementia-free life expectancy and dementia-free life expectancy as a proportion of life expectancy are given in Table 1 for ages 65 through 95 by 6-year age groups, for men and women and for both Aquitaine and France as a whole. As for disability-free life expectancy, it can be seen that at each age women have a higher dementia-free life expectancy. The results suggest that the gains in life years made by women will not be only years of dementia, nor will they be all years of good mental health; rather, women will experience the same mixture of good and poor mental health as men.

As noted above, current mortality conditions in Aquitaine appear to reflect the situation in France as a whole in about

3 years' time. If this is indeed the case then it may be predicted that there will be a future tendency in France toward an increase in dementia-free life expectancy at age 65, but that dementia-free life expectancy as a proportion of life expectancy will fall slightly (assuming dementia prevalence rates remain relatively stable).

Discussion

The calculation of dementia-free life expectancy represents a synthesis of progress made over the last 10 years in four fundamental areas of health science: (1) the elaboration of the concept of health to include quality of life and the impact of disease on society; (2) improvements in the diagnostic algorithms relating to dimensional chronic disabilities; (3) the development and validation of research instruments suitable for population research; and (4) the construction of synthetic indicators for the description of the global health status of populations.

From this study we are able to appreciate the global burden of senile dementia in Aquitaine in 1989/90, and probably for the whole of France in the coming years, given that dementia prevalence rates are not likely to differ significantly between Aquitaine and the rest of France. As the data provided by PAQUID and PAVIN closely resemble the general model of dementia prevalence derived from our previous meta-analysis of 13 recent studies, we can probably assume that the dementia-free life expectancy values found in this analysis will be approximately the same in other developed countries with similar mortality rates. The present analysis applies only to moderate and severe dementia. In public health terms, this is perhaps not too great a loss, as persons with mild dementia have only occasional difficulties that do not prevent them from living independently and the need for health service intervention for this group is negligible.

Conclusions

Like many other health expectancy studies, ours has of course been limited by the nature of the data available. As PAQUID and PAVIN are longitudinal studies, it will be possible in future years to calculate the computed period prevalence of dementia by which a fictitious cohort may be subject to the morbidity conditions occurring between consecutive rounds of the study (that is, incidence by age). As Brouard and Robine[19] have pointed out, this would provide a far more accurate estimation. The more recent calculation methods described in the literature have largely been developed to take into account the transience of disability and the multiplicity of outcomes possible. In the case of senile dementia this need is less pressing, as senile dementia is irreversible and no specific medical treatment is presently available. Furthermore, by directly combining disease prevalence with mortality, Sullivan's method implicitly takes into account the reversibility of health states.

As PAQUID and PAVIN also provide data on other aspects of health status (in particular activities of daily living and social support), future calculations may also provide estimates of life expectancy without social isolation and without loss of activities of daily living. Such estimates will permit the examination of their interrelationship with dementia-free life expectancy. Future calculations may also take possible etiological factors into account; for example, it may be possible to calculate dementia-free life expectancy for persons with different levels of education or with exposure to different risk factors. It is hoped that the present study will be used by other workers in the field of health evaluation to produce estimates of dementia-free life expectancy, thus providing a basis for international comparisons. □

Acknowledgments

The authors wish to thank Michel Isnard, INSEE, for providing them with recent and as yet unpublished mortality data for use in the calculations reported here.

References

1. Robine J-M, Ritchie K. Healthy life expectancy: evaluation of a global indicator of change in population health. *BMJ*. 1991; 302:457–460.
2. Network on Health Expectancy (REVES). *Health Expectancy Statistical World Yearbook*. Retrospective 1991 issue, Supplement to Health Expectancy Bibliography series no. 2 (Robine J-M, Romieu I, eds.). Montpellier, France: INSERM; 1991.
3. Ritchie K, Kildea D, Robine J-M. The relationship between age and the prevalence of senile dementia: a meta-analysis of recent data. *Int J Epidemiol*. 1992;21:763–769.
4. Kramer M. The rising pandemic of mental disorders and associated chronic diseases and disabilities. *Acta Psychiatr Scand*. 1980;62(suppl 285):282–297.
5. Barclay LL, Zemcov A, Blass JP, McDowell FH. Factors associated with duration of survival in Alzheimer's disease. *Biol Psychiatry*. 1985;20:86–93.
6. Molsa PK, Marttila RJ, Rinne UK. Survival and cause of death in Alzheimer's disease and multi-infarct dementia. *Acta Neurol Scand*. 1986;74:103–107.
7. Ritchie K, Colvez A, Ankri J, Ledésert B, Gardent H, Fontaine A. The evaluation of long-term care for the dementing elderly: a comparative study of hospital and collective non-medical care in France. *Int J Geriatr Psychiatry*. 1992;7:549–557.
8. Isaacs B. Geriatric patients—do their families care? *BMJ*. 1971;4:282–286.
9. Ritchie K. Life-expectancy without cognitive deterioration. In: Robine J-M, Blanchet M, Dowd JE, eds. *Health Expectancy*. London, England: Her Majesty's Stationery Office; 1992:81–86.
10. *Diagnostic and Statistical Manual of Mental Disorders*. 3d ed., revised. Washington, DC: American Psychiatric Association; 1987.
11. *Mental Disorders: Glossary and Guide to their Classification in Accordance with the Ninth Revision of the International Classification of Diseases*. Geneva, Switzerland: World Health Organization; 1978.
12. Bergmann K, Kay DWK, Foster EM, McKechnie AA, Roth M. A follow-up study of randomly selected community residents to assess the effects of chronic brain syndrome and cerebrovascular disease. In: *New Prospects in the Study of Mental Disorders in Old Age*. Amsterdam, the Netherlands: Excerpta Medica; 1971.
13. Kaneko Z. Care in Japan. In: Howells JG, ed. *Modern Perspectives in the Psychiatry of Old Age*. New York, NY: Brunner/Mazel; 1975.
14. Dartigues J-F, Gagnon M, Barberger-Gateau P, et al. The PAQUID epidemiological program on brain ageing. *Neuroepidemiology*. 1992;11:14–18.
15. Couet C, Court Y. *La Situation Démographique en 1990*. Paris, France: INSEE Résultats-Démographie Société; 1992.
16. Isnard M. *Données de Démographie Régionale 1990*. Paris, France: INSEE Résultats-Démographie Société. In press.
17. Isnard M. L'Espérance de vie selon les régions en 1989 et 1990. *INSEE Première*. 1992 (May); no. 197.
18. Sullivan DF. A single index of mortality and morbidity. *Health Serv Mental Health Admin Health Rep*. 1971;86:347–354.
19. Brouard N, Robine J-M. A method for calculation of health expectancy applied to longitudinal surveys of the elderly in France. In: Robine J-M, Blanchet M, Dowd EJ, eds. *Health Expectancy*. London, England: Her Majesty's Stationery Office; 1992:87–98.

The Prevalence and Demographics of Pathological Gamblers: Implications for Public Health

ABSTRACT

Objectives. A study of pathological gambling in five states provides information needed to address the public health threat that the expanding availability of legalized gambling poses to at-risk groups in the general population.

Methods. Over the course of this project, epidemiological data were collected to determine the prevalence of probable pathological gambling in the general population in each study state and demographic data were collected from pathological gamblers entering treatment programs in each state.

Results. Among the states surveyed, the availability of and involvement in gambling differ significantly, as does the prevalence of pathological gambling. Despite these differences, the demographics of pathological gamblers in these states are similar. Like those in the general population, pathological gamblers entering treatment in each state are similar. However, pathological gamblers entering treatment do not represent the full spectrum of individuals in the general population who experience gambling-related problems.

Conclusions. These findings raise a number of issues, including the potential impacts of continued gambling legalization on the overall rate of gambling problems in the general population and on specific at-risk groups, including women, minorities, and children. They thus have implications for policy and program decisions now being made throughout the United States. (*Am J Public Health.* 1994;84:237–241)

Rachel A. Volberg, PhD

Introduction

In the 1970s, as states experienced increasingly serious financial problems, state legislatures around the country began to legalize many types of gambling. Initially, states legalized lotteries with daily and weekly drawings and, later, instant scratch-off games. Beginning in the 1980s, states legalized pull-tab games and card rooms. Toward the end of the decade, increasing numbers of states legalized riverboats, low-stakes casinos, and video lottery terminals. With the passage of the federal Indian Gaming Regulatory Act in 1988, states around the country have seen the establishment of casino gambling on Native American lands. In 1991, Americans legally wagered $304 billion on all types of gambling, including parimutuels, lotteries, casinos, bookmaking, cards, bingo, and charitable gaming.[1]

The field of public health encompasses a myriad of environmental and behavioral threats to physical and mental health. In addressing such threats, the development of effective preventive, therapeutic, and rehabilitative interventions requires the accurate identification of the risks associated with these threats.[3,4] The proliferation of legalized gambling in the United States constitutes a public health issue because of the adverse effects that new forms of gambling can have on overall prevalence rates as well as on at-risk groups in the general population. At one end of a continuum of problematic gambling involvement is pathological gambling, characterized by a loss of control over gambling, the chasing of losses, lies and deception, family and job disruption, financial bailouts, and illegal acts.[2] In spite of recent increases in public awareness of pathological gambling as a treatable disorder and the increased availability of treatment services for individuals with gambling-related problems, this issue has yet to be conceptualized in meaningful public health terms.

Until recently, little empirical data have been available to address these issues. However, a recently completed study of the prevalence and treatment of pathological gambling in five states provides a foundation on which to begin addressing essential public health issues related to the expanding availability of legalized gambling. During this project, prevalence surveys of gambling involvement and pathological gambling were carried out in each state. In addition, information was collected on the demographic characteristics of clients in each state's gambling treatment programs. The results of those prevalence surveys are outlined below, the demographics of pathological gamblers in the general population and in treatment are reviewed, and the public health implications of these data are addressed.

Methods

Two major types of data were collected over the course of this project: (1) epidemiological data to determine the prevalence of probable pathological gambling in the general population in each study state, and (2) demographic data from pathological gamblers entering treatment programs in each state.

Prevalence Surveys

Like an earlier survey conducted in New York, the surveys funded by the Na-

Rachel A. Volberg is with Gemini Research, Northampton, Mass.

Requests for reprints should be sent to Rachel A. Volberg, PhD, Gemini Research, 80 Barrett St, No. 2, Northampton, MA 01060.

This paper was accepted May 25, 1993.

TABLE 1—Gambling Involvement and Prevalence of Pathological Gambling in the General Population

State	Lifetime Gambling Participation, % (n = 4500)	Mean No. of Gambling Activities (n = 4500)	Prevalence Rate, % (n = 4500)	1987 Per Capita Lottery Sales, $[a]
Massachusetts	90	3.51	2.3	216
Maryland	89	3.66	1.5	168
New Jersey	92	4.05	1.4	146
California	89	3.86	1.2	50
Iowa	84	3.05	0.1	35

[a]Adapted from Clotfelter and Cook.[18]

tional Institute of Mental Health were based on the South Oaks Gambling Screen (SOGS),[5,6] a 20-item scale derived from the psychiatric criteria for pathological gambling.[2] In developing the SOGS, a large pool of variables was subjected to discriminant analysis, the results of which were cross-tabulated with assessments of independent counselors. This instrument has been found valid and reliable in distinguishing probable pathological gamblers among hospital workers, university students, high school students, prison inmates, and inpatients in alcohol and substance abuse treatment programs.[7-9] The term "probable" is used to distinguish between this method of assessment and a diagnosis based on clinical evaluation.

The questionnaire for the surveys assessed respondents' (1) experience with 10 types of gambling (lotteries, casino table games, gaming machines, bingo, card games, dice games, parimutuel wagering, stockmarket activities, games of skill, and sports), (2) gambling-related problems, and (3) demographic characteristics. The surveys were carried out by telephone between April 1988 and April 1990. The number of interviews completed in each state was determined by balancing available resources, confidence intervals, and the size of each state's population. In California, 1250 interviews were completed; 1000 interviews were completed in New Jersey; and 750 were completed in Maryland, Massachusetts, and Iowa.

The sampling design in all the surveys was constructed to ensure that inferences could be drawn between the samples and the population aged 18 and over in these states. The samples from each state were stratified to proportionally represent county populations on the basis of 1980 census figures. Random-digit dialing and random selection of respondents within households were used. The refusal rate was 24% in Iowa, 27% in California, 31% in Massachusetts, 34% in Maryland, and 36% in New Jersey.

Client Demographics

In developing this project, one of the criteria used for selecting the study states was whether these states provided public funding for the treatment of pathological gambling. In 1987, New Jersey operated three outpatient treatment programs, Iowa provided outpatient services at 11 existing alcohol and substance abuse treatment centers, Maryland provided services at a psychiatric hospital and three associated community treatment programs, and Massachusetts had two hospital-based outpatient programs for pathological gamblers. California did not have treatment services for pathological gamblers but was included because of pending legislation to establish funding for such services. This legislation was subsequently defeated.

Treatment programs in all the study states based their diagnosis of pathological gambling on the psychiatric criteria that were also used in developing the SOGS.[2] Aggregate data for all clients entering publicly funded treatment programs in New Jersey and Iowa were provided by the New Jersey Division of Alcoholism and by the Iowa Department of Human Services.[10,11] In Maryland, demographic data were abstracted from client records by staff of the psychiatric hospital overseeing treatment for the state's pathological gamblers.[12] The Harvard Center for Addiction Studies provided demographic data on all clients entering treatment in Massachusetts.[13] In California, demographic data on pathological gamblers in treatment were abstracted by the author from the case records of clients treated by private practitioners. Comparison of these data with demographic information from Gamblers Anonymous[14,15] and from treatment programs in other states[16,17] indicate that the pathological gamblers entering treatment in this study are representative of pathological gamblers entering self-help and professional treatment programs throughout the United States.

Results

The study states differ in gambling involvement and in pathological gambling prevalence rates. Individuals entering treatment for gambling-related problems, as well as those in the general population, differ in gambling involvement and demographic characteristics.

Gambling Availability, Involvement, and Prevalence of Pathological Gambling

Lotteries in Massachusetts, Maryland, and New Jersey were started between 1971 and 1973 while lotteries in California and Iowa were started in 1985. Casinos were legalized in New Jersey in 1976 but have been available to California residents since the 1930s. Horse and dog racing have been available for at least 30 years in all the states, and off-track wagering has been available in the East Coast states and in California for nearly 20 years. Charitable gaming such as bingo is widely available in all the study states.

Table 1 shows differences in lifetime participation in gambling and in mean number of lifetime gambling activities for all respondents across the study states. Lifetime participation and mean number of gambling activities are highest in New Jersey and lowest in Iowa. Differences between Iowa and the other study states in these measures are statistically significant ($P < .01$, as determined by chi-square analysis). In addition, per capita lottery sales are highest in Massachusetts and lowest in Iowa.

Per capita wagering, apart from lottery sales, is difficult to correlate with prevalence rates of pathological gambling. One reason is that information on per capita wagering is not easily available on a state-by-state basis. Additionally, differences in gambling expenditures may be the result of educational and income differences that affect participation in gambling, of sociocultural differences that lead to differential participation in many types of gambling, and of the increasingly regional nature of many types of gambling.

Consistent with prior uses of the SOGS,[5,6,19] respondents in the general population surveys scoring 5 or more points were classified as "probable pathological gamblers." The prevalence rate of

probable pathological gambling in each state is also shown in Table 1.

One striking difference that emerges from these data is that the prevalence rate of probable pathological gambling in the Eastern states and California is significantly higher than that in Iowa ($P < .01$, as determined by chi-square analysis). As noted above, legalized gambling, particularly state lotteries and casinos, has been available to the residents of the East Coast states and California for many more years than it has been in the Midwest. Gambling participation and per capita lottery sales are also significantly higher in the East Coast states and California than in Iowa.

Comparing Pathological Gamblers in the General Population and in Treatment

There are significant differences, again as determined by chi-square analysis, between the overall survey sample and those respondents who scored as probable pathological gamblers along several demographic dimensions. However, the only significant difference between probable pathological gamblers from different states is that those from the East Coast states and California are significantly more likely to be non-White than those from Iowa.

Differences between the overall sample and those respondents who scored as probable pathological gamblers are detailed in Table 2. As this table makes clear, probable pathological gamblers are significantly more likely than the general population to be male and non-White, to have lower education, and to be unmarried. Probable pathological gamblers also differ from the general population in their lifetime participation in many types of gambling, particularly wagering on cards, horse and dog races, games of skill, dice games, and sports, as well as in the frequency of their wagering. Involvement in gambling was not a basis for categorizing respondents as probable pathological gamblers.

Like the characteristics of probable pathological gamblers in the general population, those of pathological gamblers entering treatment in New Jersey, Maryland, Massachusetts, Iowa, and California are strikingly homogeneous. Table 2 shows the demographic profile of pathological gamblers entering professional treatment programs in each study state as well as of those in the general population.

As this table shows, the great majority of pathological gamblers entering treatment are White, middle-aged men. By contrast, those scoring as probable pathological gamblers in the general population are more likely to be women and minorities, as well as less likely to have graduated from high school than pathological gamblers entering treatment in every state. Additionally, they are less likely to be married than those entering treatment in every state except California. It is worth noting that none of the probable pathological gamblers in the general population had ever sought treatment for a gambling problem.

TABLE 2—Characteristics of Pathological Gamblers in Professional Treatment Programs and in the General Population

	Total Survey Sample (n = 4442)	Probable Pathological Gamblers in Survey Sample (n = 58)	New Jersey (n = 155)	Maryland (n = 276)	Massachusetts (n = 137)	Iowa (n = 135)	California (n = 71)
Male, %	43	76*	93	91	93	86	93
White, %	80	64*	90	89	93	92	94
High school graduate, %	90	79*	89	84	93	87	98
Not married, %	46	62*	41	60	29	49	69
Median age, y	...	34	38	38	37	35	33

*Statistically significant difference between overall sample and probable pathological gamblers ($P < .01$) tested by chi-square analysis.

Discussion

These findings have implications for policy and program decisions now being made throughout the United States. The data raise a number of issues, including the potential impacts of continued gambling legalization on the overall rate of gambling problems in the general population and on specific at-risk groups, including women, minorities, and children.

Gambling Legalization and Prevalence Rates

In states where legal gambling has been available for less than 10 years, less than 0.5% of the adult population were classified as probable pathological gamblers. In states where legal gambling has been available for more than 20 years, approximately 1.5% of the adult population were classified as probable pathological gamblers. Together, these data support the long-standing contention of treatment professionals and researchers that increasing the availability of gambling will contribute to an increase in the prevalence of gambling-related problems in the general population.[20,21]

Research from Australia and Canada[22,23] suggests that weekly gambling and regular heavy losses are correlated with the development of gambling-related problems. We have shown that there are significant differences in lifetime gambling participation, mean numbers of gambling activities, and per capita lottery sales across the study states. While the relationship between gambling involvement, gambling expenditures, and pathological gambling across these states is not completely linear, the strong association between these variables deserves continued investigation.

Because the availability of many types of gambling is expected to increase throughout the 1990s, the public health issue is to find ways to minimize the harm that some groups will experience in relation to the expansion of gambling. Increased and reliable funding for prevention, treatment, and rehabilitation of pathological gamblers is an important first step. State legislatures may want to consider additional measures, such as establishing hotline and crisis intervention services; training mental health and substance abuse treatment professionals to recognize and address gambling-related problems among their clients; training criminal justice and prison officials to recognize the illegal activities and suicidal tendencies of some pathological gamblers; and funding research on gambling and gambling-related problems among women, minorities, and youth.

Gambling and Women

While pathological gambling affects men more often than it does women, female pathological gamblers are much less likely to enter treatment for a gambling-related problem. This is analogous to the treatment of alcoholism, in which the stigma of being a female alcoholic was

once so great that women were unlikely to seek treatment and, if they did, likely to be misdiagnosed.[24] To reach women experiencing problems related to their gambling, it will be important to focus public education and outreach efforts on venues where women are more likely to gamble, such as lottery retail outlets and bingo halls. It will also be essential to train mental health and substance abuse treatment professionals to recognize gambling problems among their women clients.

Gambling and Minorities

The data presented here show that ethnic minorities are seriously underrepresented among pathological gamblers entering treatment. An especially important issue in this regard is the potential impact of increased legal gambling on Native Americans. The Indian Gaming Regulatory Act of 1988 has led to the establishment of casino gambling on numerous reservations throughout the United States. Given the history and experience of Native Americans with alcoholism,[25–27] we must wonder whether gambling will prove equally devastating.

Preliminary information on gambling involvement and gambling-related problems among Native Americans in one state[28] shows that the prevalence rate of probable pathological gambling is significantly higher in this specific population than it is in the general population of the same state. Native American tribal governments as well as the Indian Health Service will want to develop initiatives to address the issue of pathological gambling on reservations, particularly in those states where compacts permitting casino gambling have been established.

Gambling and Youth

In Iowa and Massachusetts, we added to the survey instrument several questions that were intended to improve our understanding of how gambling problems develop. The questions focused on the respondents' age and preferred types of gambling at different points in time. Analysis shows that 8% of the Iowa respondents and 7% of the Massachusetts respondents who gambled began wagering before the age of 15. In contrast, 23% of the problem and pathological gamblers in Iowa and 36% of such gamblers in Massachusetts began wagering before the age of 15.

These data confirm clinical evidence that extensive childhood involvement in gambling is predictive of later gambling problems. Access to gaming venues by young people is difficult to control, as evidenced by the numbers of high school students and underage college students who gamble in casinos, buy lottery tickets, and place bets on horse and dog races.[18,29,30] Regulating access to new gaming venues by children and teenagers may be a critical factor in minimizing the prevalence of gambling-related problems in later life. It will also be important to increase awareness of the risks of early gambling involvement among teachers, parents, and children themselves.

Conclusion

Until well into the 20th century, excessive gambling losses were regarded as an individual failing rather than as a social or public health problem. However, data from surveys of gambling involvement in the general population and from treatment programs for pathological gambling raise serious public health concerns in relation to the proliferation of legalized gambling. As legalized gambling spreads, pressures will increase on the gaming industries, state governments, and mental health and substance abuse treatment professionals to develop effective responses to problems related to the expanding availability of gambling.

Researchers, treatment professionals, gaming industry representatives, and policymakers must work together to address these issues and to develop innovative approaches for helping individuals who experience severe problems when they gamble. It is incumbent on all these parties to consider how best to prevent the probable increases in pathological gambling that loom in the future and to ensure that services are available to those individuals whose lives are disrupted by gambling. □

Acknowledgments

This work was funded by research grant MH 44295 from the Violence and Traumatic Stress Research Branch of the National Institute of Mental Health and by the Iowa Department of Human Services.

Early versions of this paper were presented at the Eighth International Conference on Gambling and Risk Taking in London, England, June 1990, and at the Fifth Annual Conference of the National Council on Problem Gambling in Duluth, Minnesota, July 1991.

The author would like to acknowledge the contribution of Henry J. Steadman, PhD, to the development of the ideas presented here.

References

1. Christiansen EM, McQueen P. 1991 gross annual wager. *Gaming and Wagering Business.* 1992;13(7):22–37.
2. *Diagnostic and Statistical Manual of Mental Disorders.* 3rd ed. Washington, DC: American Psychiatric Association; 1980.
3. Institute of Medicine. *The Future of Public Health.* Washington, DC: National Academy Press; 1988.
4. Kessler II, Levin ML, eds. *The Community as an Epidemiologic Laboratory.* Baltimore, Md: Johns Hopkins Press; 1970.
5. Lesieur HR, Blume SB. The South Oaks Gambling Screen (SOGS): a new instrument for the identification of pathological gamblers. *Am J Psychiatry.* 1987;144: 1184–1188.
6. Volberg RA, Steadman HJ. Refining prevalence estimates of pathological gambling. *Am J Psychiatry.* 1988;145:502–505.
7. Lesieur HR, Blume SB, Zoppa RM. Alcoholism, drug abuse, and gambling. *Alcoholism: Clin Exp Res.* 1986;10:33–38.
8. Lesieur HR, Klein R. *Prisoners, Gambling and Crime.* Paper presented at the annual meeting of the Academy of Criminal Justice Sciences; March 1985; Las Vegas, Nev.
9. Lesieur HR, Klein R. Pathological gambling among high school students. *Addictive Behav.* 1987;12:129–135.
10. Weinbaum D. *Characteristics of Compulsive Gamblers Entering Treatment in New Jersey.* Trenton, NJ: New Jersey Division of Alcoholism and Drug Abuse; 1989.
11. Overland J. *Demographics of Clients Entering Treatment in Iowa, 1991.* Des Moines, Ia: Iowa Department of Human Services; 1991.
12. Ciarrocchi J, Richardson R. Profile of compulsive gamblers in treatment: update and comparisons. *J Gambling Behav.* 1989;5: 53–65.
13. Gambino B. *Demographics of Patients at the Center for Problem Gambling.* Cambridge, Mass: Harvard Center for Addiction Studies; 1990.
14. Custer RL, Custer LF. *Characteristics of the Recovering Compulsive Gambler: A Survey of 150 Members of Gamblers Anonymous.* Paper presented at the Fourth National Conference on Gambling; December 1978, Reno, Nev.
15. Nora R. *Profile Survey on Pathological Gambling.* Paper presented at the Sixth National Conference on Gambling and Risk Taking; June 1984; Reno, Nev.
16. Blackman S, Simone RV, Thoms DR, Blackman S. The gamblers treatment clinic of St. Vincent's North Richmond Community Mental Health Center: characteristics of the clients and outcome of treatment. *Int J Addict.* 1989;24:29–37.
17. Russo AM, Taber JI, McCormick RA, Ramirez LF. An outcome study of an inpatient treatment program for pathological gamblers. *Hosp Community Psychiatry.* 1984;35:823–827.
18. Clotfelter CT, Cook PJ. *Selling Hope: State Lotteries in America.* Cambridge, Mass: Harvard University Press; 1989: table 2.3.
19. Frank ML. *Casino Gambling and College Students: Three Sequential Years of Data.* Paper presented at the Third National Conference on Gambling Behavior; May 1988; New York, NY.
20. Kallick M, Suits D, Dielman T, Hybels J. *A Survey of American Gambling Attitudes and Behavior.* Research Report Series,

Survey Research Center, Institute for Social Research. Ann Arbor, Mich: University of Michigan Press; 1979.
21. Rosecrance J. *Gambling without Guilt: The Legitimation of an American Pastime*. Belmont, Calif: Wadsworth; 1988.
22. Dickerson MG. A preliminary exploration of a two-stage methodology in the assessment of the extent and degree of gambling-related problems in the Australian population. In: Eadington WR, Cornelius J, eds. *Gambling Behavior and Problem Gambling*. Reno, Nev: University of Nevada; 1993:347–363.
23. Ladouceur R, Gaboury A, Dumont M, Rochette P. Gambling: relationship between the frequency of wins and irrational thinking. *J Psychol*. 1988;122:409–414.
24. Sandmaier M. Alcoholics invisible: the ordeal of the female alcoholic. *Soc Policy*. 1980;10:25–30.
25. *Suicide, Homicide and Alcoholism among American Indians: Guidelines for Help*. Washington, DC: National Institute of Mental Health; 1973.
26. Leland J. *North American Indian Drinking and Alcohol Abuse*. Reno, Nev: University of Nevada, Desert Research Institute; 1977.
27. Andre JM. *American Indian Drinking Patterns*. Rockville, Md: Indian Health Service; 1979.
28. Volberg RA. *Gambling Involvement and Problem Gambling among Native Americans in North Dakota*. Report to the North Dakota Department of Human Services; 1993.
29. Arcuri AF, Lester D, Smith FO. Shaping adolescent gambling behavior. *Adolescence*. 1985;20:935–938.
30. Jacobs DF. Illegal and undocumented: a review of teenage gambling and the plight of children of problem gamblers in America. In: Shaffer HJ, Stein SA, Gambino B, Cummings TN, eds. *Compulsive Gambling: Theory, Research, and Practice*. Boston, Mass: Lexington Books; 1989.

WFPHA International Congress on Health, Economics, and Development

The World Federation of Public Health Associations (WFPHA) will hold its Seventh International Congress, "Health, Economics, and Development: Working Together for Change," December 5 through 8, 1994, in Bali, Indonesia. [Note revised date for conference.] Join health practitioners, policymakers, administrators, development workers, economists, researchers, and many others from governments, international organizations, and private groups in this exciting and ambitious event. Examine the linkages among health, economics, and development. Develop an action agenda to integrate health concerns with economic and development policies.

The WFPHA is a nongovernmental organization composed of national public health associations from 41 countries around the world. WFPHA congresses are held every 3 years and are cosponsored by the World Health Organization and the United Nations Children's Fund. This congress will be hosted by the Indonesian Public Health Association. Participation is open to all. The official language of the congress is English.

For further information and abstract forms, contact WFPHA Secretariat, c/o American Public Health Association, 1015 15th St NW, Suite 300, Washington, DC 20005 USA; fax (202) 789-5681. The deadline for abstract submission is *May 15, 1994*.

Sedatives and Hypnotics in Stockholm: Social Factors and Kinds of Use

Göran Blennow, MD, Anders Romelsjö, MD, PhD, Håkan Leifman, BA, Anders Leifman, MSD, and Gunilla Karlsson, MD

ABSTRACT

Objectives. The aims of the study were (1) to estimate prevalence rates of current, regular, and long-term use of sedatives and hypnotics and the incidence of regular use in an urban population and (2) to study the association between such use of drugs and sociodemographic factors, symptoms of disease, and alcohol consumption.

Methods. Data on drug use in a random sample of 6217 adults in Stockholm County were analyzed with logistic regression.

Results. The prevalence rate for current use of sedatives or hypnotics was 12.8% among men and 18.6% among women; the rate for regular use was 3.7% among males and 4.7% among females. The odds ratio for current use increased with age and was higher among unemployed persons and disability pensioners, high consumers of alcohol, persons with an increased level of symptoms, and widows. More than 25% of the persons who had used sedatives or hypnotics during the previous 2 weeks were regular users 6 months later. For persons aged 25 through 64 years, the annual incidence rate was 1.8% among men and 2.7% among women.

Conclusions. The comparatively low incidence and high prevalence of regular use implies that long-term use of sedatives and hypnotics is common. (*Am J Public Health.* 1994;84: 242–246)

Introduction

The use of sedative and hypnotic drugs has increased in Europe since World War II and peaked in the early 1970s.[1] In Sweden and elsewhere, benzodiazepines are the most frequently used drugs,[2] at least in part because of their beneficial clinical effects and low toxicity.[1-3] In a 1981 study including the United States and 10 Western European countries, Balter found a prevalence rate of 8.6% in Sweden for use of sedative drugs, once or repeatedly; the prevalence rate for the United States was estimated at 12.9%.[4] However, several reports have stressed a risk of dependence in long-term benzodiazepine maintenance.[5-7]

Studies on the use of sedative and hypnotic drugs have revealed higher than average prevalence rates among divorced, separated, or widowed persons and elderly persons,[8-10] and low rates among the best educated.[11] It has also been shown that concomitant use of alcohol and sedatives or hypnotics is common.[12,13] Thus, there seem to be associations between social factors, alcohol use, and use of sedatives and hypnotics. However, it is not known whether there are any differences in these associations between current and regular users. This is an important issue in an analysis of the risk of dependence, and it carries implications about the question of proper vs unnecessary use.

In the present study we considered the following questions:

1. What is the prevalence of current use, regular use, and long-term use, and what is the incidence of regular use?
2. Is the social pattern and the connection to alcohol consumption different for current users and regular users?

Methods

The Health of the Population Study[14] was a cross-sectional health survey of random samples of 450 inhabitants aged 18 years and older in each of the catchment areas of 14 primary care teams at four primary health care districts in Stockholm County, which has a total of 1.5 million inhabitants. The proportion of blue-collar workers in the population in the study area was 36%; the proportion of subjects aged 65 years and older was 16%; and the proportion of unmarried persons was 37%. For Stockholm County as a whole, these proportions were 32%, 19%, and 33%, respectively.[15] Our analyses are based on a net sample of 6217 subjects (with due corrections for deaths, moves, etc.). A total of 4094 subjects (65.9%) responded to a mail questionnaire followed by up to three reminders, in April through June 1984. The second reminder to main immigrant groups included a questionnaire translated into their respective native tongues. A health examination (directed to the whole sample group) was conducted between September 1984 and March 1985 by district nurses and nurses' assistants specially trained for the purpose. The average time between the investigators' receipt of the completed questionnaire and the health examination was about 6 months

The authors are with the Department of Social Medicine, Karolinska Institute, Sundbyberg, Sweden. Göran Blennow and Gunilla Karlsson are also with the Department of Psychiatry, St. Göran's Hospital, Stockholm, Sweden. Håkan Leifman is also with the Department of Sociology, Stockholm University.

Requests for reprints should be sent to Anders Romelsjö, MD, PhD, Karolinska Institute, Department of Social Medicine, Kronan Health Centre, S-172 83 Sundbyberg, Sweden.

This paper was accepted July 26, 1993.

(mean time, 172 days; range, 116–228 days). Altogether, 3801 subjects (61.4%) attended the health examination; 2652 subjects aged 18 through 64 years participated in both the questionnaire study and the health examination.

Use of Sedatives and Hypnotics

The questionnaire contained questions about use during the previous 2 weeks of "drugs . . . for daytime use (i.e., Librium, Valium, Stesolid, Apozepam, and Sobril)" (sedatives) and "drugs . . . for nighttime use" (hypnotics). The reply alternatives were "No," "Yes, occasionally," "Yes, daily or almost daily," and "Do not know." In the health examination, subjects aged 18 through 64 years were asked about sedatives or hypnotics they had used at least 3 days a week during the previous 3 months. Since the name of the drug was stated, a separate group of benzodiazepine users could be created. The questionnaire data thus allowed us to estimate prevalence rates of current use, while the health examination formed the basis for the estimation of regular use of sedatives and hypnotics—and, separately, benzodiazepines—among persons aged 18 through 64 years. The prevalence rate of long-term use, that is, use of sedatives or hypnotics indicated in both investigations, was also calculated, as well as the incidence rate (the rate for those who stated that they used sedative or hypnotic drugs at the health examination but who did not report such use in the questionnaire).

Socioeconomic Status

Socioeconomic status was analyzed for all persons younger than 65 years, which is the retirement age in Sweden. The unemployed formed one category and the disability pensioners formed another. The remaining persons (approximately 90%) who stated an occupation were classified into various groups according to the Swedish socioeconomic classification system used in censuses as a measure of social class; the system is based on the type and extent of education generally needed for various occupations.[16] In the statistical analyses, the blue-collar workers were chosen as the reference category.

Marital Status

The marital status categories were unmarried (never married), married (reference group), divorced, and widowed.

Measures of Alcohol Consumption

The reported customary consumption of liquor, fortified wine, table wine, strong beer (5.5% alcohol by volume) and ordinary beer (3.5% alcohol by volume) was converted to grams 100% ethanol, using data on the average content of the various alcoholic beverages in 1984 (Göran Christensen, the Alcohol Retail Monopoly, personal communication, February 1984). The cut-off point for high consumption was 35 g per day for men and 25 g for women, consistent with recent recommendations.[17–19]

Symptoms

A dichotomized measure was constructed from the answers to questions about frequency of symptoms ("Never," "Seldom," "Sometimes," "Often," scored 1, 2, 3, and 4, respectively) covering 22 somatic, psychological, or psychosomatic entities. These symptoms were difficulty in sleeping, difficulty in falling asleep, poor appetite, heartburn, diarrhea, constipation, epigastralgia, flatulence, headache, palpitations, breathing problems, vertigo, depression, restlessness, anxiety, tiredness, indolence, muscle tension, breast pain, arthralgia, coughing, and incontinence. The measure for the 20% of subjects with the highest score (43–76) was coded as 1, and the measure for the others was coded as 0.

Analyses

The prevalence rates for current and regular use of sedatives and hypnotics were calculated first. Then multivariate logistic regression analyses were conducted for subjects aged 25 through 64 years for whom information about these two measures was available. The age category 25 through 44 years was used as the reference category. (Because there were few subjects aged 18 through 24 years, some of whom were not employed, this age group was excluded from the multivariate analyses.) The odds ratio, with 95% confidence intervals, was used as a measure of relative risk.[20]

Results

Current Use

Use of sedatives, hypnotics, or both during the previous 2 weeks was more common among women than men (the overall rate was 16%), increased with age, and was similar in all socioeconomic groups. The prevalence rates were lower for sedative use than for the use of hypnotics (Table 1). Drug use was most common among subjects in the following categories: disability pensioners of both sexes, unemployed women, self-employed men, and divorced or widowed persons. Prevalence rates were generally low among white-collar employees at medium and high levels. Subjects with a high level of symptoms had markedly higher prevalence rates than did those with few symptoms; high consumers of alcohol more frequently reported use of drugs than did others. Only three men and seven women aged 18 through 24 years had used sedative or hypnotic drugs during the previous 2 weeks, and only three subjects in this age group were regular users.

Regular Use

The prevalence rates for use of sedatives or hypnotics at least 3 days a week for a period of at least 3 months were about 25% as high as the rates for use during the 2 weeks before the subject answered the questionnaire (Tables 1 and 2). About 75% of the regular users took benzodiazepines. The socioeconomic distribution was generally similar for regular use and for current use, with the exception of self-employed men and unmarried persons; both categories had lower rates of regular use. There were no users among widowed men and few among self-employed men; however, there were few subjects in these categories.

The average weekly alcohol consumption among current sedative or hypnotic drug users aged 25 through 64 years was 200 g of 100% alcohol for men and 72 g for women, compared with 109 g and 49 g, respectively, among nonusers.

Long-Term Use

The prevalence rate of long-term use (6 months or more) of sedatives or hypnotics among current users aged 25 through 64 years was 26% (24 of 92) for men and 28% (44 of 156) for women. If only benzodiazepines were considered, the proportions were 16% and 18%, respectively. Of the regular users, 32 men (75%) and 44 women (76%) were long-term users. Of those who responded to the questionnaire, 2.6% of the men and 4.0% of the women admitted long-term use. The socioeconomic pattern of long-term use was similar to the pattern of both current and regular use. The annual incidence rate of regular use was 1.8% for men and 2.7% for women.

Multivariate Logistic Regression Analyses

Table 3 shows clear associations between certain social factors and use of sedatives or hypnotics; in some instances,

TABLE 1—Prevalence (%) of Use of Sedatives or Hypnotics during the 2 Weeks before Receipt of Mail Questionnaire

	Men (n = 1856)			Women (n = 2238)		
	Sedatives	Hypnotics	Either	Sedatives	Hypnotics	Either
Age, y						
18–24	0.0	1.7	1.7	1.6	1.9	2.7
25–44	3.9	5.5	7.1	5.4	6.5	8.9
45–64	7.7	13.9	16.0	10.7	18.9	22.2
65–74	11.4	23.5	28.8	16.4	35.0	39.8
75+	18.0	38.1	40.3	20.3	48.5	52.1
Work status						
Disability pension[a]	26.4	38.9	44.4	25.0	33.8	40.6
Unemployed[a]	6.3	9.4	12.5	16.7	16.0	28.0
Blue-collar[a]	3.4	6.5	7.2	4.6	6.7	8.7
White-collar, low level[a]	3.5	4.9	5.7	7.0	11.4	13.6
White-collar, middle/high levels[a]	2.6	5.6	7.2	4.6	8.0	10.3
Self-employed[a]	10.1	11.5	16.5	11.4	4.6	13.6
Marital status						
Unmarried	3.9	7.1	8.0	4.8	8.4	10.2
Married	6.2	11.9	14.8	7.5	13.3	16.3
Divorced	10.3	13.4	16.1	13.1	21.0	23.4
Widowed	11.8	35.3	36.1	19.0	40.3	46.6
Symptoms						
High level	16.4	30.0	35.4	21.1	29.0	35.7
Low level	2.7	5.6	6.7	3.4	7.8	9.4
Alcohol consumption						
High	8.4	21.1	22.2	18.6	20.5	27.3
Not high	5.6	9.7	11.9	8.1	15.4	18.2
Mean	5.9	10.7	12.8	8.6	15.7	18.6

[a]Aged 18 through 64 years.

TABLE 2—Prevalence (%) of Use of Sedatives or Hypnotics at Least Three Times per Week during the 3 Months before Health Examination

	Men (n = 1354)		Women (n = 1658)	
	All Sedatives/ Hypnotics	Benzo-diazepines	All Sedatives/ Hypnotics	Benzo-diazepines
Age, y				
18–24	0.7	0.7	0.5	0.0
25–44	3.4	1.8	3.3	2.0
45–64	5.1	3.4	8.7	5.3
Work status				
Disability pension	14.3	12.9	21.6	14.4
Unemployed	9.5	0.0	12.0	8.0
Blue-collar	3.0	1.4	3.1	1.9
White-collar, low level	4.3	3.2	5.8	3.3
White-collar, middle/high levels	2.5	0.8	1.2	0.6
Self-employed	2.5	2.5	0.0	0.0
Marital status				
Unmarried	4.2	1.9	3.2	1.0
Married	3.0	1.9	3.9	2.4
Divorced	9.3	6.5	8.5	6.1
Widowed	0.0	0.0	12.3	8.8
Symptoms				
High level	0.0	0.0	27.6	18.5
Low level	2.9	1.8	3.8	2.0
Alcohol consumption				
High	4.7	1.9	7.8	5.9
Not high	3.2	1.9	4.6	2.6
Mean	3.7	2.2	4.7	2.8

Note. Data are for subjects aged 18 through 64 years.

these associations are statistically significant. For current use among men, the odds ratios were significantly increased among the unemployed, disability pensioners, the self-employed, high alcohol consumers, and those with various symptoms. For current use among women, the odds ratios were significantly increased among those aged 45 through 54 years, low-level white-collar employees, widows, high consumers of alcohol, and those with various symptoms. In separate analyses of current users (excluding persons who were also regular users), a similar pattern was found. The most pronounced difference was the markedly increased odds ratio (4.41) for men with high alcohol consumption (95% confidence interval = 2.28, 8.55).

Men who were unemployed, disability pensioners, unmarried, or divorced or who showed symptoms had statistically significantly increased odds ratios for regular use of sedatives and hypnotics (Table 3). Among women, the odds ratios were statistically significantly increased for those aged 55 through 64 years and for those with symptoms. Thus we found both similarities and differences by gender, as well as differences in the sociodemographic pattern for regular use compared with current use. For men, odds ratios for being unmarried or divorced were statistically significantly increased only for regular use, and odds ratios for high alcohol consumption were increased only for current use. For women, the odds ratios were not increased for those who were unmarried or divorced with current or regular use of sedatives or hypnotics.

Discussion

The prevalence rate of current use of sedatives or hypnotics was 12.8% for men and 18.6% for women; these rates are higher than those found in some other surveys, including one from Sweden.[21] One interview survey of a random sample of the population indicated a prevalence rate of 3.6% for use of sedatives and 5% for hypnotics "occasionally or more frequently" during a previous 14-day period in Stockholm County in 1988[22]; another interview survey found a prevalence rate of 5.6% for use of sedatives and 5.4% for hypnotics.[21] Both surveys had an upper age limit of 74 years. If we apply this upper age limit to the Health of the Population Study sample, the prevalence rate of sedative use is reduced to 6.7% (5.4% for men and 7.7% for women) and that of hypnotic use is reduced to 11.2% (9.7% for men and

13% for women), and the prevalence rates are 11.7% for men and 15.8% for women for use of either type of drug. The higher prevalence rates of current use, especially of hypnotics, among those who responded to the questionnaire could be due partly to a reduced reluctance to admit such use in a questionnaire compared with a personal interview. It has been claimed that sensitive items usually attain a higher prevalence rate in questionnaires than in interviews.[23] It is therefore possible that underreporting contributed to a reduction of the prevalence rates found at the health examination. Another explanation could be that the use of sedatives and hypnotics actually was higher in the investigated areas in Stockholm County than in other areas.

We lack information about the use of sedatives and hypnotics among persons who did not participate in the study. However, because the prevalence in the Health of the Population Study sample for inpatient care with a diagnosis of alcoholism was two to three times higher among the nonparticipants than among the participants,[19] and because we demonstrated an association between high alcohol consumption and current use of sedatives or hypnotics in the present study, we believe that the prevalence of such use is also higher among the nonparticipants.

The increase in prevalence of use with age is in agreement with observations made by others.[24–27] When approximately 20% of persons aged 75 years and older use sedatives and approximately 40% use hypnotics, it is questionable whether appropriate indications are always present. Results of other studies indicate that use of sedatives and hypnotics implies an increased risk for falls and other accidents.[28]

A considerable variation in use of sedatives and hypnotics with socioeconomic status[13] and education[29] has been reported. We found a greatly increased prevalence of use among disability pensioners and unemployed persons of both sexes (Table 3). Being unemployed or on a disability pension is sometimes not just a short-term problem; it may have a long-term impact, especially on disability pensioners, who may feel they will never again take part in the working life. It is reasonable to believe that being unemployed or on a disability pension may be associated with a number of psychological problems caused by, for instance, social isolation. However, drug use in itself may be the reason for unemployment or a disability pension. There were increased odds ratios for current use among female low-level white-collar employees, who constitute the second largest category of gainfully employed women. One possible partial explanation is that a considerable number of women with limited work experience outside the home entered the work force; some may have received inappropriate, monotonous, or stressful jobs, developed (psychosomatic) symptoms, and subsequently been prescribed sedatives or hypnotics by physicians. We found increased odds ratios for regular use among divorced and unmarried men and for current use among widows, findings that have been reported elsewhere.[29] Other studies have shown that divorced persons have higher levels of psychic distress.[7,29] Widowed persons are more likely to suffer from depression.[30] These factors may contribute to increased odds ratios, but they do not explain the sex differences.

TABLE 3—Association between Sociodemographic Factors and Current and Regular Use of Sedatives or Hypnotics

	Current Use[a]		Regular Use[b]	
	Men	Women	Men	Women
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Age, y				
45–54	1.29 (0.75, 2.19)	1.17 (0.74, 1.85)	2.33 (0.38, 3.70)	1.96 (1.22, 7.36)
55–64	1.27 (0.73, 2.19)	2.07 (1.37, 3.13)	1.18 (0.83, 6.57)	2.99 (0.77, 4.97)
Work status				
Unemployed/disability pension	6.10 (3.20, 11.61)	3.25 (1.83, 5.77)	7.42 (2.02, 27.24)	2.98 (1.05, 8.48)
White-collar, low level	1.26 (0.67, 2.36)	1.58 (1.04, 2.41)	1.38 (0.38, 4.98)	1.66 (0.71, 3.90)
White-collar, middle/high level	1.56 (0.93, 2.63)	1.50 (0.93, 2.43)	1.20 (0.38, 3.76)	0.36 (0.08, 1.73)
Self-employed	3.33 (1.61, 6.89)	2.00 (0.79, 5.09)	1.51 (0.29, 7.84)	...
Marital status				
Unmarried	0.78 (0.48, 1.26)	0.94 (0.61, 1.45)	3.49 (1.33, 9.15)	0.82 (0.28, 2.38)
Divorced	1.14 (0.62, 2.10)	1.48 (0.96, 2.28)	4.09 (1.25, 13.40)	1.21 (0.48, 3.04)
Widowed	1.26 (0.67, 2.36)	5.61 (3.24, 9.72)	...	...
High alcohol consumption	2.14 (1.24, 3.70)	2.31 (1.17, 4.55)	0.75 (0.21, 2.74)	1.19 (0.24, 5.83)
With symptoms	7.30 (4.76, 11.18)	4.72 (3.37, 6.62)	5.97 (2.57, 13.86)	7.56 (3.59, 15.94)

Note. Data are from a multiple logistic regression analysis of data for subjects aged 25 through 64 years; the age group 25 through 44 years served as the reference category. OR = odds ratio; CI = confidence interval.
[a]Use during the previous 2 weeks.
[b]Use at least 3 days a week for at least 3 months.

The markedly increased odds ratio for current, but not regular, use of sedatives or hypnotics in subjects of both sexes with high alcohol consumption may indicate that some high consumers occasionally use these drugs as an alternative to alcohol to achieve relaxation and sleep, maybe as a consequence of heavy drinking.

The distinction between current and regular use is somewhat arbitrary. No information is available on use prior to the period covered by the questionnaire or in between the two measurement points. It may very well be that some of the current users also used sedatives or hypnotics regularly or in periods preceding the presentation of the questionnaire.

The prevalence rate of regular use (at least 3 days per week during the previous 3 months), 3.7% in men and 4.7% in women, is about one-fourth the rate of current use. At the health examination about 6 months after the questionnaire was administered, 27% of the current users were still using the drugs, that is, were long-term users. Our measure of regular use is a clear indication of long-term use, inasmuch as 75% of the regular users were long-term users. In a prospective study in Jämtland County, 15% of the persons who were prescribed benzodiazepines for the first time in 1976 were regular users 10 years later (G. Boethius, MD, oral communication, October 1989), and 1.6% of men and 2.9% of women were regular users of psychoactive drugs in an analysis of four random samples of the Swedish population from 1975 to 1981.[31] As the follow-up period was longer in the former study and the latter excluded hypnotics, the present results seem to be roughly in concordance with the results of these two studies and with those of studies from other countries in which investigators

found that 10% to 18% of benzodiazepine users continued their use for 4 months or more.[32,33]

The annual incidence rates of regular use (1.8% for men and 2.7% for women) might be considered fairly modest. The much higher prevalence rates constitute an important indication that many people use these drugs for long periods of time, contrary to recommendations.[3] It should be kept in mind in this context that the incidence rates are based on a limited number of new cases and consequently have a low degree of precision.

The multivariate analyses showed that some of the independent factors related to current use were the same for both regular and long-term users, a finding that increases the credibility of the results.

The marked association between a high level of symptoms and current and regular use of sedatives or hypnotics among persons of both sexes in the working age groups is in agreement with the findings of other studies[8,14,29] and may be regarded as a reflection of a rational use of these drugs. Conversely, it might be argued that the high proportion of users with symptoms indicates insufficient effects of the sedative and hypnotic drugs. Some persons in this category should probably receive other treatment in addition to or instead of sedatives or hypnotics, and others might do as well or better without sedatives and hypnotics. It is regrettable that we do not know what proportion of users did not actually improve, got worse, or did not change after they started to use sedative and hypnotic drugs. This information is a prerequisite for a judgment of the extent of improper use.

Since there are hardly any documented therapeutic long-term effects of benzodiazepines[34]—the dominant subgroup of sedatives and hypnotics whose use can lead to dependence—and with the well-established risk of dependence in long-term use, the proportion of 27% long-term users among current users in this study is an alarmingly high figure. It indicates that a substantial minority of current "temporary" users continue to use sedatives or hypnotics for longer periods than are suitable for effective sedative or hypnotic results. Certain general principles for prescribing sedatives and hypnotics[35] can be suggested: (1) Use nonpharmacological means of promoting sleep whenever possible and reassure patients that 5 to 6 hours' sleep per night is normal. (2) Avoid prescribing benzodiazepines—especially for the elderly—and try to withdraw benzodiazepines as soon as possible. (3) As a general rule, prescribe sedatives and hypnotics only for short periods. (4) Warn patients of the risk of dependence.

The physician should also consider the important social factors associated with use of sedatives and hypnotics—apart from age and symptoms—and should be aware that temporary use of such drugs may be an indicator of at least transitory high alcohol consumption. □

References

1. Allgulander C. History and current status of sedative-hypnotic drug use and abuse. *Acta Psychiatr Scand.* 1986;73:465–478.
2. *Svensk Läkemedelsstatistik* [Swedish Drug Statistics]. Stockholm, Sweden: Apoteksbolaget; 1991.
3. Lader MH, Petursson H. Rational use of anxiolytic/sedative drugs. *Drugs.* 1983;25:514–528.
4. Balter MB, Mannheimer DI, Mellinger GD, Uhlenhuth EH. A cross-national comparison of anti-anxiety/sedative drug use. *Curr Med Res Opin.* 1984;8:5–20.
5. Petursson H, Lader MH. Benzodiazepine dependence. *Br J Addict.* 1981;76:133–145.
6. Lader MH. Dependence on benzodiazepines. *J Clin Psychiatry.* 1983;44:121–127.
7. Owen RT, Tyrer P. Benzodiazepine dependence: a review of the evidence. *Drugs.* 1983;25:385–398.
8. Cottler LB, Robins LN. The prevalence and characteristics of psychoactive medication use in a general population study. *Psychopharmacol Bull.* 1983;19:746–751.
9. Dunbar GC, Perera MH, Jenner FA. Pattern of benzodiazepine use in Great Britain as measured by a general population survey. *Br J Psychiatry.* 1989;155:836–841.
10. Ashton H, Golding JF. Tranquilizers, predictors and possible consequences. Data from a large United Kingdom survey. *Br J Addict.* 1989;84:541–546.
11. Marinier RL, Pihl RO, Wilford C, Lapp JE. Psychotropic drug use by women: demographic, lifestyle, and personality correlates. *Drug Intell Clin Pharmacy.* 1985;19:40–45.
12. Ross HE, Glaser FB, Stiasny S. Sex differences in the prevalence of psychiatric disorders in patients with alcohol and drug problems. *Br J Addict.* 1988;83:1179–1192.
13. Genebadia J, Blay-Pueyo C, Soler-Vila M. Risk factors in the use of benzodiazepines. *Fam Pract.* 1988;5:283–288.
14. Boström G, Diderichsen F, Nordström G. *The Health of the Population Study in Stockholm County.* [in Swedish]. Sundbyberg, Sweden: Karolinska Institute; 1987. Report, blue series no. 146.
15. Boström G, Diderichsen F, Nordström G. *The Health of the Population Study in Stockholm County: Methods and Analyses of Nonparticipation* [in Swedish]. Sundbyberg, Sweden: Karolinska Institute; 1987. Report, green series no. 171.
16. *Swedish Socioeconomic Classification* [in Swedish]. Stockholm, Sweden: Statistics Sweden; 1982.
17. Hollstedt C, Rydberg U. Hazardous alcohol consumption and early diagnosis of alcohol-linked disease [in Swedish, summary in English]. *Läkartidningen.* 1981;78:795–799.
18. *That's the Limit: A Guide to Sensible Drinking.* London, England: Health Education Council Publications; 1983.
19. Romelsjö A. The relationship between alcohol consumption and social status in Stockholm. Has the pattern of alcohol consumption changed? *Int J Epidemiol.* 1981;18:842–851.
20. Rothman KJ. *Modern Epidemiology.* Boston, Mass: Little, Brown & Co; 1986.
21. *Variabler och koder för LNU81* [Variables and Codes for the Survey of Living Levels]. Stockholm, Sweden: Swedish Institute for Social Research; 1984.
22. Kühlhorn E, Leifman H. Befolkningens bruk av alkohol och psykofarmaka i empirisk belysning [The use of psychotropics and alcohol in the population in an empirical analysis]. *Social Forskning.* In press.
23. Hochstim J. A critical comparison of three strategies of collecting data from households. *J Am Stat Assoc.* 1967;62:976–989.
24. Ashton H. Benzodiazepine withdrawal: an unfinished story. *BMJ.* 1984;288:1135–1140.
25. Petursson H, Lader M. Withdrawal from long-term benzodiazepines treatment. *BMJ.* 1981;283:634–635.
26. Tyrer P, Owen R, Dawling S. Gradual withdrawal of diazepam after long-term therapy. *Lancet.* 1983;i:1402–1406.
27. Woods J, Katz JL, Winger G. Abuse liability of benzodiazepines. *Pharmacol Rev.* 1987;39:251–413.
28. Oster G, Huse D, Adams S, Imbimbo J, Russel M. Benzodiazepine tranquilizers and the risk of accidental injury. *Am J Public Health.* 1990;80:1467–1470.
29. Swartz M, Landerman R, George LK, Melville ML, Blazer D, Smith K. Benzodiazepine anti-anxiety agents: Prevalence and correlates of use in a southern community. *Am J Public Health.* 1991;81:592–596.
30. Mor V, McHorney C, Sherwood S. Secondary morbidity among the recently bereaved. *Am J Psychiatry.* 1986;143(2):158–163.
31. Allgulander C. Psychoactive drug use in a general population sample, Sweden: correlates with perceived health, psychiatric diagnosis, and mortality in an automated record-linkage study. *Am J Public Health.* 1989;79:1006–1010.
32. Mant A, Duncan-Jones P, Saltman D, et al. Development of long-term use of psychotropic drugs by general practice patients. *BMJ.* 1988;296:251–254.
33. Williams P. Long-term benzodiazepine use in general practice. In: Freeman H, Rue Y, eds. *The Benzodiazepines in Current Clinical Practice.* London, England: Royal Society of Medicine; 1987:19–31.
34. Committee of the Review of Medicines. Systematic review of the benzodiazepines. *BMJ.* 1980;280:910–912.
35. Hecker R, Burr M, Newbury G. Risk of benzodiazepine dependence resulting from hospital admission. *Drug Alcohol Rev.* 1992;11:131–135.

Societal-Level Predictors of Groups' Drinking Patterns: A Research Synthesis from the Collaborative Alcohol-Related Longitudinal Project

Kaye Middleton Fillmore, PhD, Jacqueline M. Golding, PhD, E. Victor Leino, PhD, Catherine R. Ager, BA, and Heidi P. Ferrer, BS, with the Collaborative Alcohol-Related Longitudinal Project

ABSTRACT

Objectives. Period effects of per capita consumption of alcohol and divorce rates are assessed for change in quantity and frequency among age/sex groups in multiple longitudinal studies.

Methods. Twenty-five studies of quantity and 29 studies of frequency are used. Studies are from 15 nations and cover periods of 1 to 21 years. Models predict the standardized mean difference for quantity and frequency based on period effects and group-level and methodological variables.

Results. When both the period effects of per capita consumption and the divorce rate are considered, the divorce rate significantly predicts change in quantity and frequency. An increase in the divorce rate is associated with a stronger decrease in frequency among younger people; men are more likely than women to decrease their frequency of drinking when divorce rates rise.

Conclusions. Multiple societal-level factors should be considered critical in influencing the drinking patterns of groups. These results suggest that an increase in the divorce rate is associated with more "dry" social contexts, characterized possibly by drinking patterns of a more "volitive" nature (i.e., heavier quantity per occasion and less frequent drinking). (*Am J Public Health.* 1994; 84:247–253)

Introduction

Understanding social factors that contribute to change in drinking practices within groups is a challenging research task. Traditionally, several designs have been used to address this issue, including historical analyses,[1] aggregate-level statistics,[2,3] and cross-sectional general population studies.[4,5]

This paper uses an alternative design to address alcohol-related social change in groups. It systematically combines results from multiple longitudinal studies of the general population and evaluates the associations of both societal and study-level characteristics with the dependent variables: the mean differences of quantity per occasion and the frequency of drinking between final measurement (T_2) and Time 1 (T_1) for age/sex groups. Societal-level statistics are the primary independent variables of theoretical interest in predicting group-level change of drinking practices.

The advantages of the cross-level design are that (1) multiple samples are used, representing different historical periods and national contexts; (2) characteristics of groups *within* their own societies/period are used as predictors in addition to societal-level predictors; (3) period effects (i.e., historical influences on the lives of individuals measured as change over the study period) in multiple social settings are systematically evaluated using societal-level factors; and (4) effects may be evaluated over variable lengths of time using samples of differing measurement intervals.

Two societal-level explanations are advanced to explain changes in group drinking:

1. The dominant theory maintains that the proportion of drinkers at high levels of intake (e.g., 40 to 80 or more g/day) is related to the average consumption level in a group (typically measured by the overall per capita consumption of alcohol in nations).[2,6,7] Our analysis addresses the association between changes in the per capita consumption of nations and shifts in the means of drinking measures among specific age/sex groups in those nations. Thus, we can determine if shifts in per capita consumption affect all groups in the society equally. To do this, we ask two questions. First, controlling for characteristics of the age/sex groups and the samples, do the Time 1 per capita consumption and the change over time (the period effect) of per capita consumption predict the mean differences in quantity per occasion and frequency per month for various age/sex groups? We expect the period effect of increasing per capita consumption to be associated with increases in the frequency and decreases in the quantity of drinking per occasion because "wet" societies have been characterized by higher frequency of drinking and more moderate quantity per occasion.[8] Second, are some age groups (and gender) more likely to be affected by the period effect than others?

2. Divorce has been linked to high-intake drinking on at least two levels of analysis: (1) cross-sectional studies of the

The authors are with the Institute for Health and Aging, Department of Social and Behavioral Sciences, University of California, San Francisco. For a list of the members of the Collaborative Alcohol-Related Longitudinal Project, see the Acknowledgments.

Requests for reprints should be sent to Kaye M. Fillmore, Institute for Health and Aging, University of California, San Francisco, Box 0646, Laurel Heights, San Francisco, CA 94143-0646.

This paper was accepted June 23, 1993.

general population report that the proportion of drinkers at high levels of intake is negatively associated with the stability of marriage on the individual level[9]; and (2) divorce rates, with other indicators of societal-level stress, have been linked to higher per capita consumption and alcohol problems.[10] Using divorce rate as a measure of societal-level stress, we again ask two questions. First, controlling for characteristics of the age/sex groups and the samples, do the Time 1 divorce rate and change in the divorce rate between measurements (the period effect) predict the mean differences in quantity and frequency for age/sex groups? Second, are some age groups (and gender) more likely to be affected by the period effect than others?

3. The final model tests the joint contribution of the Time 1 measures and the period effects of divorce and per capita consumption in predicting the mean changes in quantity and frequency.

The paper is based on the idea that societal-level changes (i.e., period effects measuring the variable at two points in time, T_2–T_1) may be implicated in changing the drinking patterns of groups in multiple societies/periods. It addresses the possibility that the period effects may be differentially distributed within the societies/periods by the age and sex of the groups studied. Finally, it tests the independent effect of each of the societal-level/period variables by controlling for the other.

Methods

Twenty-five studies from the archive of the Collaborative Alcohol-Related Longitudinal Project[11] were used to measure the quantity of drinking per typical occasion; 29 studies were used to measure the frequency of drinking per month. These studies represent sample sizes ranging from 535 to 12 686. Samples consist of abstainers and drinkers at Time 1. Studies are from Canada, Czechoslovakia, Denmark, the Federal Republic of Germany, Finland, Ireland, Israel, New Zealand, Norway, Poland, the Shetland Islands, Sweden, Switzerland, the United Kingdom, and the United States; periods of measurement range from 1960 to 1986 at Time 1. Intervals between measurements range from 1 to 21 years. After studies were disaggregated by age of first measurement in 5-year increments between ages 9 and 39, and in 10-year increments to 80 years and older, effect estimates from samples were calculated for designated age/sex categories yielding a base (n) of 141 for the frequency analysis and of 134 for the quantity analysis.

Basically, the models tested herein are ones in which changes in drinking between two points in time among designated age/sex groups from the different studies are the objects of inquiry. These changes are the effect sizes. The primary variables thought to predict these changes are the period effects of per capita consumption of alcohol and divorce rates over the same period in which the changes in drinking were observed in the various studies. Other variables thought to contribute to these changes are the age and gender of the groups and the characteristics of the study in which the change was measured. These measures and the model are detailed below.

The dependent variables are the standardized mean differences of frequency of alcohol consumption and of quantity per occasion between final measurement and Time 1, both recoded to consistent levels of measurement across the studies to reflect current drinking in the last month.

Demographic variables hypothesized to influence the distribution of effect estimates are the age of the group at Time 1 and the sex of the group (0 = female; 1 = male). Study characteristics are the interval in years between measurements and the percentage of attrition in the sample between Time 1 and final measurement.

Societal-level variables are (1) per capita consumption of alcohol in the nation at Time 1 in US gallons of absolute alcohol (T_1PCC), (2) the difference in per capita consumption between final measurement and Time 1 (D-PCC), (3) the interaction of age of the group at Time 1 and per capita consumption at Time 1 (T_1Age × T_1PCC), (4) the interaction of gender of the group and per capita consumption at Time 1 (Sex × T_1PCC), (5) the interaction of age at Time 1 and the difference in per capita consumption between final measurement and Time 1 (T_1Age × D-PCC), (6) the interaction of gender and the difference in per capita consumption between final measurement and Time 1 (Sex × D-PCC), (7) the Time 1 divorce rate per 1000 (T_1Div), (8) the difference in the divorce rate between final measurement and Time 1 (D-Div), (9) the interaction of age at Time 1 and the Time 1 divorce rate (T_1Age × T_1Div), (10) the interaction of gender and the Time 1 divorce rate (Sex × T_1Div), (11) the interaction of age at Time 1 and the difference in the divorce rate between final measurement and Time 1 (T_1Age × D-Div), and (12) the interaction of gender and the difference in the divorce rate between final measurement and Time 1 (Sex × D-Div). The natural logs were taken of T_1PCC and T_2PCC, and the difference was formed (D-PCC). The interaction terms are included to test the hypotheses that the associations of the societal variables at Time 1 or their period effects with changes in quantity or frequency differ for age groups and gender.

The unit of analysis used for comparison across studies is the effect size, or the standardized difference between mean frequency of alcohol consumption at first and final measurement points.[12] The standardized mean difference indicates the degree to which (and the direction in which) age/sex groups are, relatively speaking, stable or changing. The effect size has been widely used in meta-analyses because of its ready interpretability as a scale-free estimate of the magnitude and direction of effects. Originally developed to compare results from treatment and control groups in experimental studies, the measure can also be used effectively in nonexperimental longitudinal research. For this purpose, the effect size constitutes a scale-invariant estimate of the magnitude and direction of group change over time, expressed in standard deviation units. All values reported here have been calculated using the variant formulation of the effect size developed by Hedges and Olkin[13] because of its superior properties as an estimate of the population effect size in small sample contexts.

Estimates of the relative impact of multiple factors on effect sizes are generated by using weighted least squares regression procedures with effect sizes as the dependent variable. Each study result is weighted by the reciprocal of the variance of the effect size, thus giving greater weight to more precise estimates.[13] Results of each model include a standardized regression coefficient (β) for each predictor and a test of the significance of its relationship to the dependent variable. The use of the standardized regression coefficient (the effect on the dependent variable, in standard deviation units, of a one–standard deviation change in an explanatory variable) should be distinguished from the use of the standardized mean differences (the effect sizes), which are the "standardized" dependent variable values. The null hypothesis for this analysis is that there is no influence of the societal independent variables on the stan-

dardized mean differences after adjusting for demographic factors and study characteristics.[12]

The correlations between the primary societal-level factors for the quantity and frequency samples (Table 1) indicate that rising per capita consumption occurs in studies where T_1PCC was originally relatively low; the correlation between T_1Div and D-Div is much more modest; and T_1Div and T_1PCC are virtually unrelated. D-Div and D-PCC are positively related, suggesting that increases in these period effects sometimes co-occur in the samples studied.

TABLE 1—Correlations between the Primary Societal-Level Variables for the Frequency of Drinking Sample and the Quantity of Drinking Sample

	Frequency of Drinking			Quantity of Drinking		
	D-PCC	T_1Div	D-Div	D-PCC	T_1Div	D-Div
T_1PCC	−.704	.227	−.352	−.690	.174	−.369
D-PCC		−.305	.529		−.289	.601
T_1Div			−.079			−.086

Note. Time 1 per capita consumption of alcohol (T_1PCC), difference in per capita consumption between final measurement and Time 1 (D-PCC), Time 1 divorce rate (T_1Div), and difference in divorce rate between final measurement and Time 1 (D-Div).

TABLE 2—Standardized Regression Coefficients (β) for Time 1 Frequency of Drinking per Month and Time 1 Quantity of Drinking per Occasion: The Effects of Age, Sex, Time 1 per Capita Consumption, and Interactions

	Time 1 Frequency of Drinking per Month			Time 1 Quantity of Drinking per Occasion		
Predictor	β	SE	*P*	β	SE	*P*
Age	.438	.068	.000	−.126	.083	.128
Sex	.418	.068	.000	.322	.081	.000
T_1PCC	.116	.069	.094	.137	.081	.092
T_1Age × T_1PCC	−.038	.069	.581	−.257	.083	.002
	Multiple R = .616; F = 20.834; P = .000; df = 136			Multiple R = .408; F = 6.443; P = .000; df = 129		
Age	.442	.068	.000	−.066	.083	.432
Sex	.416	.068	.000	.305	.084	.000
T_1PCC	.121	.068	.077	.126	.084	.136
Sex × T_1PCC	.003	.068	.962	−.019	.084	.822
	Multiple R = .615; F = 20.712; P = .000; df = 136			Multiple R = .323; F = 3.764; P = .006; df = 129		

Note. T_1PCC = Time 1 per capita consumption of alcohol; T_1Age × T_1PCC = the interaction of age at Time 1 and per capita consumption of alcohol at Time 1; Sex × T_1PCC = the interaction of gender and Time 1 per capita consumption of alcohol.

Results

Per Capita Consumption of Alcohol

Table 2 displays the cross-sectional relationships of frequency/quantity at Time 1 with age of the group, sex of the group, T_1PCC, and T_1Age × T_1PCC in the first two models. Sex × T_1PCC is substituted for T_1Age × T_1PCC in the second two models.

Age is not significantly associated with quantity and is positively associated with frequency, suggesting that older groups drink more often. Men show higher frequency and quantity in the aggregate than women do.

T_1PCC is not significantly associated with quantity or frequency although the associations are approaching significance and are positive. The interaction of T_1Age × T_1PCC is not associated with frequency in a statistically significant way but is negatively associated with quantity. Models calculated for separate age groups show that young people in higher per capita consumption contexts are more likely to drink more heavily on each occasion than are young people in lower per capita consumption contexts, and quantity per occasion among those in their middle and later years is not related to per capita consumption context. The interaction of Sex × T_1PCC is not significantly related statistically to either frequency or quantity.

To assess the associations of the period effects of D-PCC with the mean change of quantity and frequency and to examine which age/sex groups, if any, are affected, the effects of age, sex, interval between measurements of the studies, sample attrition, T_1PCC, the period effects of D-PCC, and T_1Age × D-PCC are simultaneously entered in the first two models (Table 3); Sex × D-PCC is substituted for T_1Age × D-PCC in the second two models.

Age is not significantly related to the mean difference of quantity but is negatively related to that of frequency. There are no significant gender differences. The interval between measurements of studies is positively related to frequency and negatively related to quantity, indicating that groups with longer intervals between measurement will show greater increases in frequency and greater decreases in quantity. Sample attrition is not significantly related to quantity but is positively related to frequency, indicating that groups with higher attrition increase in mean frequency over time.

There are no main effects of T_1PCC on the mean change of quantity or frequency. There is no main effect of D-PCC on the mean difference of frequency but there is a positive (albeit marginal) effect on the mean difference of quantity, suggesting that quantity per occasion rises with the period effect of D-PCC. (Noteworthy, however, is that the correlation of D-PCC and the mean difference of quantity is close to zero.) The relationships of the interactions of T_1Age × D-PCC and Sex × D-PCC with the mean difference of either quantity or frequency are not significant.

The Divorce Rate

Table 4 displays the cross-sectional relationships of frequency/quantity at Time 1 with age of the group, sex of the group, T_1Div, and T_1Age × T_1Div in the first two models. Sex × T_1Div is substituted for T_1Age × T_1Div in the second two models.

T_1Div is positively associated with T_1 frequency and T_1 quantity (the latter being only marginally significant), suggesting that higher societal-level stress is associated with both heavier and more frequent drinking in the cross-section. The interac-

TABLE 3—Standardized Regression Coefficients (β) for Mean Difference of Frequency of Drinking per Month and Quantity of Drinking per Occasion: The Difference in Per Capita Consumption of Alcohol and Interactions

Predictor	Mean Difference of Frequency of Drinking per Month β	SE	P	Mean Difference of Quantity of Drinking per Occasion β	SE	P
Age	−.527	.076	.000	−.072	.110	.516
Sex	.093	.068	.178	.070	.081	.390
Interval	.242	.087	.006	−.457	.104	.000
Attrition	.328	.101	.001	−.133	.124	.284
T_1PCC	−.012	.102	.903	.0002	.133	.998
D-PCC	.131	.118	.143	.352	.185	.059
T_1Age × D-PCC	.160	.104	.125	.001	.107	.991
	Multiple R = .647; F = 13.653; P = .000; df = 133			Multiple R = .447; F = 4.462; P = .000; df = 125		
Age	−.508	.076	.000	−.075	.110	.493
Sex	.109	.080	.178	.103	.124	.408
Interval	.299	.079	.000	−.459	.096	.000
Attrition	.289	.098	.004	−.133	.122	.278
T_1PCC	−.034	.102	.739	.004	.128	.975
D-PCC	.007	.124	.953	.354	.168	.038
Sex × D-PCC	.029	.079	.717	.043	.123	.725
	Multiple R = .639; F = 13.110; P = .000; df = 133			Multiple R = .448; F = 4.482; P = .000; df = 125		

Note. T_1PCC = Time 1 per capita consumption of alcohol; D-PCC = the difference in per capita consumption between final and first measurement; T_1Age × D-PCC = the interaction of age at Time 1 and the difference in per capita consumption between final and first measurement; Sex × D-PCC = the interaction of sex and the difference in per capita consumption between final and first measurement.

TABLE 4—Standardized Regression Coefficients (β) for Time 1 Frequency of Drinking per Month and Time 1 Quantity of Drinking per Occasion: The Effects of Age, Sex, Time 1 Divorce Rate, and Interactions

Predictor	Time 1 Frequency of Drinking per Month β	SE	P	Time 1 Quantity of Drinking per Occasion β	SE	P
Age	.423	.066	.000	−.048	.084	.566
Sex	.443	.066	.000	.315	.084	.000
T_1Div	.244	.068	.005	.184	.087	.060
T_1Age × T_1Div	−.021	.068	.756	−.027	.086	.755
	Multiple R = .652; F = 25.187; P = .000; df = 136			Multiple R = .343; F = 4.312; P = .003; df = 129		
Age	.420	.065	.000	−.044	.084	.595
Sex	.441	.065	.000	.315	.083	.000
T_1Div	.266	.067	.000	.167	.085	.053
Sex × T_1Div	.083	.066	.211	−.025	.084	.770
	Multiple R = .657; F = 25.833; P = .000; df = 136			Multiple R = .343; F = 4.309; P = .003; df = 129		

Note. T_1Div = Time 1 divorce rate; T_1Age × T_1Div = the interaction of age at Time 1 and the Time 1 divorce rate; Sex × T_1Div = the interaction of gender and the Time 1 per capita consumption of alcohol.

tions of either age or gender with T_1Div are not significant.

Assessment of the period effect (Table 5) simultaneously enters the effects of age, sex, interval between measurements of the studies, sample attrition, T_1Div, D-Div, and the interaction of T_1Age × D-Div in the first two models; the interaction of Sex × D-Div is substituted for T_1Age × D-Div in the second two models.

T_1Div is negatively related to the mean difference of quantity (marginal) and frequency, suggesting that groups originally in "high" divorce rate contexts will decrease and groups in "low" divorce rate contexts will increase their quantity and frequency over time. D-Div is positively associated with the mean difference of quantity and negatively associated with the mean difference of frequency, suggesting that when divorce rates rise, quantity per occasion will rise and frequency will fall.

T_1Age × D-Div is not statistically significant for the mean difference of quantity but is marginally significant for frequency, indicating that frequency is likely to fall among all age groups with a rising divorce rate but that the association is stronger among the young. (Separate models for age groups [not shown] demonstrate this finding.) The interaction of gender and the period effect of the divorce rate is not associated with quantity but is negatively associated with frequency. Separate models for each sex (not shown) reveal that while both men and women show decreases in frequency when the divorce rate increases, the effect is stronger and significant among men and is not significant among women.

The Independent Societal-Level Period Effects of per Capita Consumption and the Divorce Rate

To determine if the period effects of either per capita consumption or the divorce rate are associated with the mean differences after controlling for the other period effect, age, sex, interval between measurements of the studies, sample attrition, T_1Div, T_1PCC, D-PCC, and D-Div are simultaneously entered (Table 6).

The period effect of per capita consumption is not significantly associated with the mean difference of quantity when the period effect of the divorce rate is in the model; however, it is marginally significant (positive) in association with the frequency of drinking per month. Noteworthy is that the sign changes: a positive correlation and a negative beta. This is therefore interpreted with caution owing to the multicolinearity among the independent variables and to the low correlation between D-PCC and the mean frequency of drinking.

The period effect of the divorce rate is negatively associated with the mean difference of frequency and positively associated with the mean difference of quantity.

Thus it can be concluded that when both societal-level effects are entered into the same model, an increase in the per capita consumption of alcohol may be implicated in changing the mean frequency

of drinking. However, the divorce rate is more likely to have a significant influence over time, indicating an elevation of the quantity of drinking and a decrease in the frequency.

Discussion

This paper assesses the period effects of per capita consumption of alcohol and divorce rates on change in quantity of drinking per occasion and on frequency of drinking per month among age/sex groups drawn from multiple longitudinal studies representing different historical periods and nations. The basic assumption in the analysis is that alcohol-related social change within groups is not random but, instead, is at least partly derived from larger societal changes.

Noteworthy is that these analyses are concerned with the mean difference of quantity per occasion and frequency per month of age/sex groups, not with the proportions of persons in these populations who are heavy and/or frequent drinkers. This being the case, large mean changes would not necessarily suggest that it is the more entrenched heavy and/or frequent drinkers who are changing but, rather, that it is larger portions of the population.

Clearly the strongest effect in the period effect models is the interval of time between measurement points (ranging from 1 to 21 years), characterizing this group of studies as increasing in frequency and decreasing in quantity as interval of time increased.

Some results both reflect and inform understanding of maturation with respect to drinking. Previous cross-sectional research has shown that age is positively related to frequency of drinking and negatively related to quantity per occasion.[14] While only approaching significance, the cross-sectional results of the interaction between age and per capita consumption at Time 1 suggest that younger people in higher per capita consumption contexts are more likely to drink more per occasion than younger people in lower per capita consumption contexts, whereas older people in both contexts are not differentiated by their quantity per occasion. This interpretation is stated with caution.

The longitudinal findings also refine what we know about maturation and drinking. While the interaction of age and the period effect of the divorce rate is only marginally significant, it suggests that the frequency of drinking falls among all age groups when divorce rates increase but that the relationship is stronger among the young. The lesser effect among older groups may be owing, in part, to the collective influence of the increased prevalence of divorced persons in these groups, which, along with the group's changing drinking patterns, our past research has shown to be related to changing individual-level drinking patterns.[15,16] While this suggests that an increase in societal-level stress might be implicated in altering the decline in frequency observed as a result of increasing age among the cohorts most likely affected, it should be remembered that the nations under scrutiny vary with respect to the socioreligious constraints on divorce (e.g., annulment and separation in some countries may be equivalent to divorce in others). But even taking this caveat into consideration, if social stress (in this case, divorce rates) increases, such a contextual effect could change the trajectory in age cohorts from what would be

TABLE 5—Standardized Regression Coefficients (β) for Mean Difference of Frequency of Drinking per Month and Quantity of Drinking per Occasion: Interval, Attrition, Divorce Rate, and Interactions

	Mean Difference of Frequency of Drinking per Month			Mean Difference of Quantity of Drinking per Occasion		
Predictor	β	SE	*P*	β	SE	*P*
Age	−.367	.062	.000	−.075	.114	.510
Sex	.104	.058	.081	.061	.081	.449
Interval	.369	.069	.000	−.382	.102	.000
Attrition	.129	.084	.127	−.022	.110	.842
T_1Div	−.354	.075	.000	−.183	.095	.057
D-Div	−.228	.087	.010	.228	.113	.046
T_1Age × D-Div	.134	.096	.119	.059	.134	.660
	Multiple R = .757; F = 25.583; P = .000; df = 133			Multiple R = .461; F = 4.821; P = .000; df = 125		
Age	−.389	.061	.000	−.105	.094	.268
Sex	−.007	.069	.920	.074	.100	.459
Interval	.438	.067	.000	−.361	.088	.000
Attrition	.121	.082	.141	−.015	.108	.893
T_1Div	−.369	.073	.000	−.187	.095	.052
D-Div	−.329	.061	.000	.196	.084	.021
Sex × D-Div	−.181	.069	.009	.023	.100	.820
	Multiple R = .766; F = 27.061; P = .000; df = 133			Multiple R = .460; F = 4.795; P = .000; df = 125		

Note. T_1Div = the divorce rate at Time 1; D-Div = the difference in the divorce rate; T_1Age × D-Div = the interaction of age at Time 1 and the difference in the divorce rate; Sex × D-Div = the interaction of gender and the difference in the divorce rate.

TABLE 6—Standardized Regression Coefficients (β) for Mean Difference of Frequency of Drinking per Month and Quantity of Drinking per Occasion: The Differences in the Divorce Rate and in per Capita Consumption

	Mean Difference of Frequency of Drinking per Month					Mean Difference of Quantity of Drinking per Occasion				
Predictor	*r*	β	SE	*t*	*P*	*r*	β	SE	*t*	*P*
Age	−.365	−.320	.069	−4.638	.000	−.142	−.106	.112	−.951	.344
Sex	.130	.084	.059	1.441	.152	.066	.063	.080	.783	.435
Interval	.337	.383	.068	5.651	.000	−.365	−.425	.101	−4.216	.000
Attrition	.362	.172	.091	1.895	.060	−.054	−.119	.124	−.958	.340
T_1Div	−.494	−.414	.079	−5.219	.000	−.186	−.106	.106	−.997	.321
T_1PCC	−.154	−.129	.088	−1.472	.143	−.192	.0002	.126	.001	.999
D-PCC	.131	−.205	.110	−1.864	.065	.032	.238	.182	1.308	.193
D-Div	−.178	−.338	.063	−5.402	.000	.132	.165	.084	1.960	.052
	Multiple R = .759; F = 22.526; P = .000; df = 132					Multiple R = .479; F = 4.624; P = .000; df = 124				

Note. T_1PCC = Time 1 per capita consumption; T_1Div = Time 1 divorce rate; D-PCC = difference in the per capita consumption rate; D-Div = difference in the divorce rate.

expected from maturation alone. Such an effect could reflect increased group cohesiveness among older persons; alternatively, it could also reflect sample attrition among this age group. However, if the change in divorce rate changes the "natural" trajectory of the frequency of drinking as people mature, special attention to this age group may be warranted.

Therefore, these two social forces, one measuring the normative relative "wetness" of a society and the other measuring an aspect of social stress, may interact with age in ways that give us insight beyond the effects of maturation alone.

One interaction of gender with the societal-level measures is significant for the period effect findings. The interaction between the period effect of the divorce rate and gender in association with the mean difference of frequency indicates that frequency significantly decreases among men but not among women in contexts of increasing divorce rates. It is shown that the main effect of an increasing divorce rate is associated with decreasing frequency and increasing quantity of consumption. This may mean that societal stress is related to the emergence of a "drier" society as evidenced by the decreased frequency of drinking among the young, particularly among men, and by the increased quantity per occasion among all groups.

The cross-sectional findings widen our understanding of the relationships of both per capita consumption of alcohol and the divorce rate to drinking. As expected, Time 1 per capita consumption is positively related to Time 1 frequency of drinking (marginal), and Time 1 divorce rate is positively related to both Time 1 frequency and quantity (marginal for quantity). This suggests that both norms of the relative wetness of societies and indicators of social stress may be useful in describing the drinking of the groups in those societies and periods studied. However, knowledge of the per capita consumption at Time 1 is of less importance in predicting the mean difference of frequency and quantity. On the other hand, the divorce rate is a relatively strong predictor of the mean difference of both quantity and frequency (in models where only divorce rates are evaluated) and of frequency only (in models including the per capita consumption measures), such that groups originally in higher Time 1 divorce rate contexts are more likely to decrease in frequency over time. The latter finding may also suggest a regression to the mean as well as the "law of initial value" (i.e., the greater the distance from the mean, the greater the change possible). Alternatively, it may suggest that societies characterized by higher social stress are more likely to be heavier and more frequent drinkers initially but, as these groups age, both heavy and frequent drinking are likely to decrease, most probably owing to maturation (a finding that has been observed in most studies and at most ages in the middle to later years of life).[14]

Finally, the impact of the period effects themselves requires comment. The period effects of changing per capita consumption on group drinking practices are not as strong as expected when using this research design. (The effect is positive on quantity for models without the divorce rate variables included and nonsignificant when those variables are included, and the effect is negative on frequency for models with the divorce rate variables included and nonsignificant when those variables are not included.) While a change in per capita consumption would be expected to mold the *volume* of drinking in these groups (both being equivalent under the ideal conditions of measurement), the mean differences do not respond consistently across models to these period effects. This calls for caution in developing societywide programs to curtail per capita consumption because it may not be as strong a variable in molding the drinking *patterns* of groups as previously thought.

When the two period effects are entered, it is the divorce rate that shows a significant negative association with the mean difference of the frequency of drinking and a positive association with the quantity of drinking, suggesting that this form of societal-level stress may alter drinking patterns by possibly transforming the groups' social contexts into "drier" ones. This suggests, but certainly does not prove, that societal-level stress may be related to the emergence of more "volitive" drinking patterns, a hypothesis worthy of future pursuit.

These findings suggest that there is merit in consideration of multiple societal-level factors as period effects in the study of social change surrounding drinking. The practical importance is that some specialists in alcohol-related prevention have suggested that intervention on the societal level, such as reducing the per capita consumption of alcohol, will reduce the average consumption in a group and hence reduce the proportion of heavy drinkers in the society. But while this may be true of the overall volume of drinking, it does not seem to be consistently the case in drinking patterns of age/sex groups. Results here suggest that a reduction in the incidence of divorce in society may be associated with a wetter social context by curbing an increase in quantity per occasion and stimulating an increase in frequency. However, if the problems of a dry social context are to be avoided, reduction in this form of societal-level stress may not be warranted. Nonetheless, it seems probable that a reduction in the divorce rate might reduce more volitive drinking (although that remains to be demonstrated) and may perhaps have more beneficial effects for a larger group of social problems. □

Acknowledgments

This work was supported by a National Institute on Alcohol Abuse and Alcoholism (NIAAA) grant (R01 AA07034) and by an NIAAA Research Scientist Development Award (K01 AA00073) to the first author.

The Collaborative Alcohol-Related Longitudinal Project is included in the plan of work of the National Institute on Alcohol Abuse and Alcoholism as a World Health Organization (WHO) Collaborating Center on Research and Training in Alcohol-Related Problems, and is also affiliated with the WHO Global Program on Prevention and Control of Alcohol and Drug Abuse. The following persons were collaborators on this paper, having reviewed it and its findings for accuracy and representation of their data and project goals: Salme Ahlstrom, PhD (Finland), Peter Allebeck, MD (Sweden), Arvid Amundsen, PhD (Norway), Jules Angst, MD (Switzerland), Gellisse Bagnall, PhD (Scotland), Ann Brunswick, PhD (USA), Sally Casswell, PhD (New Zealand), Lorraine de Labry, PhD (USA), Norman Giesbrecht, PhD (Canada), Bridget Grant, PhD (USA), Thomas Greenfield, PhD (USA), Joel Grube, PhD (USA), Bernd Guether, PhD (Federal Republic of Germany), Thomas Harford, PhD (USA), Ludek Kubicka, PhD (Czechoslovakia), Steve Manske, PhD (Canada), Mark Morgan, PhD (Ireland), Harold Mulford, PhD (USA), Leif Ojesjo, MD (Sweden), David Peck, PhD (Scotland), Martin Plant, PhD (Scotland), Chris Power, PhD (UK), Bruce Ritson (UK), Lee Robins, PhD (USA), Anders Romelsjo, MD (Sweden), David Rosen, MD (USA and Shetland Islands), Ronald Schlegel, PhD (Canada), Martin Sieber, PhD (Switzerland), Rainer Silbereisen, PhD (Federal Republic of Germany), Ronald Stall, PhD (USA), Meir Teichman, PhD (Israel), Richard Wilsnack, PhD (USA), and Sharon Wilsnack, PhD (USA).

References

1. Levine H. *Demon of the Middle Class: Self-Control and the Ideology of Temperance in 19th Century America.* Berkeley, Calif: University of California; 1978. PhD dissertation.
2. Ledermann S. *Alcool, alcoolisme, alcoolisation. Institut National d'Etudes Demographiques.* Paris, France: Presses Universitaires de France; 1956.
3. Holder HD, Blose JO. Impact of changes in

distilled spirits availability on apparent consumption: a time series analysis of liquor-by-the-drink. *Br J Addict.* 1987;82:623–631.

4. Hingson R, Heeren T, Scotch N. Effects of Maine's .02 law to reduce teenage drunk drinking. Presented at the Alcohol Epidemiology Section, International Council on Alcohol and Addictions; June 1985; Rome, Italy.
5. Kendell RE, de Roumanie M, Ritson EB. Effect of economic changes on Scottish drinking habits 1978–82. *Br J Addict.* 1983; 78:365–379.
6. Makela K, Room R, Single E, et al. *Alcohol, Society and the State: A Comparative Study of Alcohol Control.* Toronto, Canada: Addiction Research Foundation; 1981.
7. Skog O-J. The wetness of drinking cultures. A key variable in epidemiology of alcoholic liver cirrhosis. *Acta Med Scand Suppl.* 1986;703:157–184.
8. Room R. Response to alcohol-related problems in an international perspective: characterizing and explaining cultural wetness and dryness. Presented at "La ricerca Italiana sulle bevande alcoliche nel confronto internazationale"; September 1989; Santo Stefano Belbo, Italy.
9. Knupfer G. The prevalence in various social groups of eight different drinking patterns, from abstaining to frequent drunkenness: analysis of 10 US surveys combined. *Br J Addict.* 1989;84:1305–1318.
10. Linsky AS, Colby JP, Straus MA. Stress, drinking culture and alcohol problems: a partial test of Bales' theory. *Soc Sci Med.* 1987;24:875–883.
11. Fillmore KM, Hartka E, Johnstone BM, et al. The Collaborative Alcohol-Related Longitudinal Project: preliminary results from a meta-analysis of multiple longitudinal studies. *Br J Addict.* 1991;86:1203–1210.
12. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. New York, NY: Academic Press; 1969.
13. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis.* New York, NY: Academic Press; 1985.
14. Fillmore KM, Hartka E, Johnstone BM, et al. A meta-analysis of life course variation in drinking. *Br J Addict.* 1991;86:1221–1268.
15. Fillmore KM, Johnstone BM, Leino EV, Ager CR. A cross-study contextual analysis of effects from individual-level drinking and group-level drinking factors: a meta-analysis of multiple longitudinal studies from the Collaborative Alcohol-Related Longitudinal Project. *J Stud Alcohol.* 1992; 54:37–47.
16. Leino EV, Ager CR, Fillmore KM, Johnstone BM. A cross-study contextual analysis of group influences and individual influences of changing marital status on individual drinking behavior: a meta-analysis of multiple longitudinal studies from the Collaborative Alcohol-Related Longitudinal Project. In preparation.

Television Beer Advertising and Drinking Knowledge, Beliefs, and Intentions among Schoolchildren

Joel W. Grube, PhD, and Lawrence Wallack, DrPH

ABSTRACT

Objectives. The relationships between television beer advertising and drinking knowledge, beliefs, and intentions were investigated in a survey of schoolchildren. The research was guided by a theoretical model specifying that awareness of advertising, and not mere exposure, is necessary for it to have an effect on beliefs or behaviors.

Methods. Participants were a random sample of 468 fifth- and sixth-grade schoolchildren from a northern California community. Data were collected in the home with a combination of self-administered questionnaires and structured interviews.

Results. Nonrecursive statistical modeling indicated that awareness of television beer advertising was related to more favorable beliefs about drinking, to greater knowledge of beer brands and slogans, and to increased intentions to drink as an adult. The effects of advertising awareness on knowledge, beliefs, and intentions were maintained when the reciprocal effects of beliefs, knowledge, and intentions on awareness were controlled.

Conclusions. The findings suggest that alcohol advertising may predispose young people to drinking. As a result, efforts to prevent drinking and drinking problems among young people should give attention to countering the potential effects of alcohol advertising. (*Am J Public Health.* 1994;84:254–259)

Introduction

Children have well-developed beliefs about alcohol and its effects even before they have personal experience with drinking.[1–4] Although parents, peers, and other environmental influences are important in shaping these beliefs and, ultimately, drinking behaviors, alcohol advertising also may be an important source through which children learn about this drug. Popular belief in the effects of alcohol advertising is widespread. A recent survey conducted by an alcohol industry group found that 73% of the general public agreed that alcohol advertising is a major contributor to underage drinking.[5] Similarly, two recent federal reports have recommended that alcohol advertising, especially that directed at young people, be limited or banned and that equal time requirements be established to provide for the dissemination of public health messages about drinking.[6,7] In addition, legislative proposals in the US Senate and House of Representatives would require health messages on alcoholic beverage advertisements in the print and electronic media. To what extent are concerns about the influence of alcohol advertising on young people well founded? The present study addressed this issue by investigating the relationships between awareness of television beer advertising and drinking knowledge, beliefs, and intentions of schoolchildren.

Most of the available research on alcohol advertising and youth has shown small but significant correlations between exposure to alcohol advertising and drinking beliefs and behaviors. In one study, exposure to alcohol advertising was found to be related to heavy drinking, problem drinking, and drinking in hazardous situations (e.g., while driving) among youth.[8] These relationships persisted when demographic variables and beliefs about drinking by peers and parents were statistically controlled. Another study showed that adolescents who were heavily exposed to advertising were more likely to agree that drinkers possess valued characteristics such as being attractive, athletic, or successful.[9] More recent research has shown that young drinkers are more likely to be exposed to alcohol advertisements, are more accurate in identifying brands of beer from edited alcohol advertisements, and are more favorably disposed toward these advertisements than are nondrinkers.[10] A few studies have found that exposure to advertising is only modestly related to consumption of alcohol by young people when significant demographic and personal variables are controlled.[11,12]

Although most studies have found a significant relationship between alcohol advertising and drinking beliefs and behaviors among young people, some issues remain. At a conceptual level, most researchers have assumed that simple exposure to alcohol advertising is sufficient to bring about attitude and behavioral changes. However, an information processing model of advertising suggests that advertisements have significant effects on knowledge, beliefs, intentions, and behaviors only when attended to and remembered.[13,14] According to this perspective, *awareness* of advertising, not simply ex-

Joel W. Grube is with the Prevention Research Center, Berkeley, Calif. Lawrence Wallack is with the School of Public Health, University of California, Berkeley.

Requests for reprints should be sent to Joel W. Grube, Prevention Research Center, 2150 Shattuck Ave, Suite 900, Berkeley, CA 94704.

This paper was accepted May 27, 1993.

Editor's Note. See related annotation by Mosher (p 180) and article by Madden and Grube (p 297) in this issue.

posure to it, is important. By focusing on simple exposure, previous studies may have underestimated the importance of advertising for the drinking beliefs and behaviors of young people. In addition, previous research has relied on correlational analyses. As a result, it provides no evidence regarding the direction of the relationship between alcohol advertising and drinking beliefs and behaviors.[15] Although alcohol advertising may influence these beliefs and behaviors, it is also possible that those who are predisposed to drink seek out information about alcohol and thus are more attentive to relevant advertising.

The present study attempted to address some of these issues through an investigation of television beer advertising and drinking knowledge, beliefs, and intentions among schoolchildren. Going beyond earlier studies, an information processing model of advertising effects was used to guide the design and analyses. In addition, nonrecursive statistical modeling techniques were used to estimate the effects of awareness of television beer advertising on drinking knowledge, beliefs, and intentions and the simultaneous effects of knowledge, beliefs, and intentions on awareness.

Methods

Procedures

Overview. Data were obtained from a sample of schoolchildren in a suburban city in northern California. The data were collected in the home by trained interviewers from the Survey Research Center of the University of California, Berkeley. Sensitive questions, such as those addressing drinking intentions, were included in a self-administered instrument. Less threatening questions such as those about television viewing were included in a face-to-face interview. One of the child's parents completed a self-administered questionnaire and answered several interview questions concerning family background and demographics. Approximately 1 hour was required to complete the interviews and surveys. A $10 compensation was offered to reimburse the parents and child for their time.

Sample. The initial sample consisted of 603 fifth and sixth graders drawn at random from a complete roster provided by the local public school district. Interviews were completed with 468 children (a response rate of 77.6%).

The final sample was nearly evenly split between boys (48.9%) and girls (51.1%). The respondents ranged in age from 10 to just under 14 years old. Their median age was 11.9 years. About 32% of the children were Caucasian, 29% were Asian American (predominantly Filipino, but also including Chinese, Japanese, and Southeast Asian), 26% were African American, 7% were Latino, 4% were Pacific Islander, and 3% were of other ethnic or racial heritages. The sample was largely working and middle class, and the median family income was $33 000. Most parents reported some education beyond high school, and mothers and fathers had both completed an average of about 13 years of education.

Measures

Television viewing. Weekday television viewing was determined by asking the children how much time they spent watching television on a typical school day before school, after school until dinner, and from dinner until bedtime. Also, four weekend viewing periods were addressed: Saturday morning until noon, Saturday noon until bedtime, Sunday morning until noon, and Sunday noon until bedtime. These items were summed into separate measures representing hours of weekday viewing, hours of weekend morning viewing, and hours of weekend afternoon and evening viewing. The children also were asked how many televised sports events they had viewed in the last 4 weeks ("none" to "three or more"). This question was included because television sports have very high rates of alcohol advertising.[16]

Awareness of beer advertising. Awareness of beer advertising was measured during the interview by showing the children five still photographs from currently televised beer commercials (Coors Light, Budweiser Light, Coors, Michelob Dry, Budweiser). Any references to product or brand were blocked out of the photographs. The children were asked whether they could remember having seen the commercial depicted in each photograph and, if so, to name the brand being advertised. These items were scored 0 (had not seen the advertisement), 1 (had seen the advertisement but could not name the brand), or 2 (had seen the advertisement and named the correct brand).

Knowledge of beer brands and slogans. Knowledge of beer brands was measured by asking the children to name as many brands of beer as they could without any prompts. The interviewer recorded up to five different brands. Next, the interviewer presented the children with a list of five beer slogans. The children were asked to match each slogan with its correct brand. A summary measure was derived by counting the number of correct matches.

Alcohol beliefs. Beliefs about the positive aspects of drinking were measured by reading the children five statements that addressed alcohol's relation to (1) friendship or sociability, (2) reward for work, and (3) relaxation. These items were chosen to reflect the main themes in beer advertisements as identified in previous research.[17] The children were asked to indicate, on 4-point scales, the extent to which they thought each item was true or false ("certain it is true" to "certain it is false"). Negative beliefs were similarly measured with five items targeting (1) birth defects, (2) drinking and driving, and (3) general health problems.

Intention to drink as an adult. Because drinking levels are very low for this age group, the primary outcome variable was intention to drink as an adult. (Because only 5% of the respondents had ever had more than a taste of an alcoholic beverage, it was not possible to investigate the effects of advertising on drinking. However, drinking intentions have been found to be related to future drinking behaviors among adolescents.) Intentions were ascertained on the self-administered questionnaire with items asking the children to indicate, on 4-point scales, how often they thought they would drink beer, wine coolers, wine, and liquor when they were grown up ("at least once a week" to "not at all").

Perceived parental and peer approval of drinking. Measures of perceived approval were obtained on the self-administered questionnaire by asking the children to indicate, on 5-point scales, how happy or unhappy their mothers, fathers, best friends, and other close friends would be if the children themselves were to drink alcohol. Scales representing parental approval ($\alpha = .84$) and peer approval ($\alpha = .79$) were derived by calculating the mean values of the relevant items.

Perceived parental and peer drinking. Perceptions of parental and peer drinking were measured in the self-administered questionnaire by having the children indicate, on 6-point scales, how often they thought their mothers, fathers, best friends, and other friends had consumed alcohol in the past year ("none" to "almost every day"). Mean scores for parental drinking ($\alpha = .62$) and peer drinking ($\alpha = .81$) were calculated.

TABLE 1—Measurement Models for Latent Variables

Measured Variable	Standardized Factor Loading	Unstandardized Factor Loading	SE	t	P
Drinking intention					
Liquor	.58	1.00	...	...	...
Beer	.81	2.37	.272	8.68	<.001
Wine coolers	.71	1.91	.235	8.11	<.001
Wine	.58	1.59	.210	7.59	<.001
Positive beliefs					
Bars are a good place to hang out	.48	1.00	...	...	...
Having a few beers is a good reward	.72	1.29	.162	7.97	<.001
Drinking makes it easier to get to know people	.57	0.92	.133	6.91	<.001
Drinking is a good way to relax	.69	1.10	.137	8.03	<.001
Good friends drink together	.64	1.22	.157	7.79	<.001
Negative beliefs					
Drinking can harm health	.53	1.00	...	...	...
Dangerous to drive after drinking	.75	1.59	.254	6.23	<.001
Mother's drinking can harm unborn child	.45	0.84	.166	5.06	<.001
Mother's drinking can cause birth defects	.39	0.84	.171	4.92	<.001
Shouldn't ride with drinking driver	.52	1.20	.214	5.59	<.001
Awareness of advertising					
Coors Light	.47	1.00	...	...	...
Budweiser Light	.40	0.95	.158	5.98	<.001
Coors	.36	0.83	.149	5.57	<.001
Michelob Dry	.34	0.77	.145	5.35	<.001
Budweiser	.33	0.92	.169	5.44	<.001

Note. The unstandardized loading of the first variable on each factor was fixed at 1 to identify the variance of the latent variables.

Demographic and background variables. Family income and child's ethnicity were obtained from the parent's self-administered questionnaire, age was asked directly of the child, and sex was recorded by the interviewer. Ethnicity was coded into four dummy variables representing Caucasian, Black, Asian, and Latino respondents. Sex was also coded as a dummy variable (0 = female, 1 = male).

Results

The relationships between awareness of advertising and drinking intentions, beliefs, and knowledge were examined through a generalized least squares, nonrecursive structural equations analysis implemented with EQS.[18,19] Because the data were nonnormal, robust estimates of the standard errors were obtained and the Satorra-Bentler scaled chi-square was used to assess the fit of the models.[20]

Latent variables were used to represent awareness of beer advertising, drinking intentions, beliefs about the positive aspects of drinking, and beliefs about the negative aspects of drinking. Table 1 provides the factor loadings from the measurement model for these latent variables. Correlations were allowed between error terms for four pairs of items in the measurement model: (1) wine and wine cooler intentions ($r = .22$, $P < .001$), (2) liquor and wine intentions ($r = .22$, $P < .001$), (3) beliefs that drinking can harm an unborn child and that drinking can cause birth defects ($r = .24$, $P < .001$), and (4) recognition of the Budweiser Light commercial and belief that drinking is a good reward after a day's work ($r = .18$, $P < .001$). (In order to ensure empirical identification of the model, an equality constraint was placed on the correlations between the errors for wine and wine cooler intentions and the errors for wine and liquor intentions. The Lagrange multiplier test indicated that this constraint was reasonable, ($\chi^2[1, n = 458] = 0.11$, $P < .75$.)

The information processing framework was translated into an initial structural model specifying unidirectional effects from awareness of beer advertising to beliefs and knowledge and from beliefs to intentions. Exogenous variables considered to be important as potential predictors of awareness, knowledge, beliefs, and intentions were included in the covariance matrix for the initial model, but no effects between these variables and the endogenous variables were specified. (The exogenous variables included in the initial model were frequency of viewing sports programs, hours of weekend morning television viewing, hours of weekend afternoon and evening television viewing, hours of weekday television viewing, perceived parental approval of drinking, perceived peer approval of drinking, perceived parental drinking, perceived peer drinking, race, age, sex, family income, and years of parents' education.) After the initial model was solved, a specification search was undertaken with Lagrange multiplier tests to ascertain which exogenous variables were related to endogenous variables. Effects were added to the model only if they also were consistent with previous research and theory.[21] Exogenous variables that were unrelated to any of the endogenous variables were dropped. Next, the paths representing the reciprocal effects of positive beliefs and knowledge on awareness of advertising and of intentions on positive beliefs were added, and the model was reestimated. (One way to uniquely identify these simultaneous effects is to specify exogenous variables that have independent direct effects on each of the endogenous variables involved in a reciprocal relationship. Because no exogenous variable uniquely predicted negative beliefs, the reciprocal paths involving this variable were not estimated.) These reciprocal effects were retained in the model even when they were nonsignificant in order to control for the potential influences of knowledge, beliefs, and intentions on awareness. Finally, correlations between the disturbance terms for knowledge of brands and positive beliefs ($r = .30$, $P < .01$) and knowledge of slogans and positive beliefs ($r = .22$, $P < .01$) were added to the model. The final structural model is depicted in Figure 1 (solid arrows represent statistically significant effects, and broken arrows represent nonsignificant effects). This model provided a good fit, scaled χ^2 (330, n = 458) = 369.88, $P < .07$. The normed and comparative fit indexes for this model were .976 and .997, respectively.

Table 2 shows the structural (regression) coefficients and significance tests from the final model. Four findings are particularly noteworthy. First, awareness of alcohol advertising was significantly re-

lated to increased knowledge of beer brands and slogans and to more positive beliefs about drinking. Second, awareness was *indirectly* related to intentions to drink as an adult through positive beliefs ($\beta = .19$, $t = 2.11$, $P < .05$). Third, awareness of advertising was not significantly related to beliefs about the negative aspects of drinking. Fourth, knowledge of beer brands and slogans increased awareness of advertising.

In addition, children who expected to drink more frequently as adults had more favorable beliefs about drinking, believed that their peers were more approving of drinking, and believed that their parents drank more frequently. They also tended to be male and had parents with more years of education. Sex also had an independent effect on intention to drink beer over and above its effect on overall drinking intention. Boys reported that they were more likely than girls to drink this beverage as an adult ($\beta = .10$, $t = 2.51$, $P < .05$).

Children who perceived more parental approval of drinking believed that drinking was more likely to have positive consequences. Also, children who perceived more peer approval of drinking thought drinking was less likely to have negative consequences. Similarly, children who believed that their peers drank more frequently knew more beer slogans, and children who perceived more peer approval for drinking named more brands of beer. Boys also knew more slogans and brands than did girls. Finally, children who were more aware of alcohol advertising also believed that their parents drank more frequently and were more frequent viewers of sports and weekend afternoon and evening television.

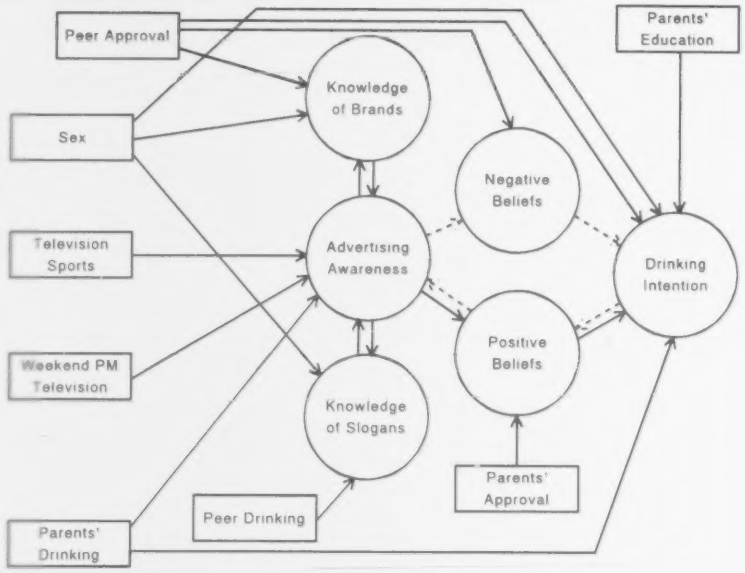


FIGURE 1—Final model of alcohol advertising and drinking knowledge, beliefs, and intentions.

TABLE 2—Structural Parameters for Final Model of Advertising Effects

Predictor	Standardized Coefficient	Unstandardized Coefficient	SE	*t*	*P*
Intention					
Negative beliefs	.04	.02	.025	0.71	...
Positive beliefs	.32	.16	.057	2.84	<.01
Peer approval	.26	.11	.024	4.60	<.001
Parental drinking	.28	.07	.014	4.75	<.001
Sex	.14	.09	.040	2.24	<.05
Parental education	.11	.02	.008	2.39	<.05
Positive beliefs					
Intention	−.07	−.13	.255	−0.52	...
Awareness of advertisements	.47	.59	.166	3.54	<.001
Parental approval	.22	.26	.074	3.55	<.001
Negative beliefs					
Awareness of advertisements	.04	.05	.098	0.53	...
Peer approval	−.13	−.12	.056	−2.09	<.05
Brands					
Awareness of advertisements	.24	.58	.220	2.65	<.01
Peer approval	.14	.24	.067	3.57	<.001
Sex	.26	.68	.126	5.37	<.001
Slogans					
Awareness of advertisements	.22	.34	.133	2.59	<.01
Peer drinking	.15	.14	.043	3.13	<.01
Sex	.17	.29	.086	3.32	<.001
Awareness of advertisements					
Positive beliefs	.17	.13	.094	1.42	...
Brands	.38	.15	.037	4.19	<.001
Slogans	.17	.11	.056	1.98	<.05
TV sports	.23	.09	.026	3.61	<.001
Weekend PM TV	.24	.03	.008	4.15	<.001
Parents' drinking	.19	.07	.022	3.34	<.001

Discussion

Children who were more aware of beer advertisements held more favorable beliefs about drinking, intended to drink more frequently as adults, and had more knowledge of beer brands and slogans. These effects were maintained when the reciprocal effects of knowledge, beliefs, and intentions on awareness were included in the model. Thus, the findings provide support for the hypothesis that awareness of alcohol advertising *influences* children's drinking beliefs, knowledge, and intentions. Interestingly, awareness of advertising was unrelated to beliefs about the negative aspects of drinking. This finding may reflect the fact that alcohol advertisements overwhelmingly associate drinking with positively valued activities and consequences such as romance, sociability, and relaxation.[17]

Significant reciprocal effects were also found. Knowledge of beer brands and slogans appeared to increase awareness of

beer advertising. These reciprocal effects may result from the fact that children who are curious about drinking seek information about alcohol. Beer commercials may represent one source of such information.

Although the present study goes beyond previous research on alcohol advertising and youth, there are two issues that should be kept in mind in evaluating the results. First, it was not possible to address drinking behaviors. Because only 5% of the children had ever had a whole drink, it was inappropriate to include this variable in the analysis. Similar research with older adolescents will be necessary to investigate the relationship between alcohol advertising awareness and drinking behaviors. Second, the findings are based on cross-sectional data. Although, when reciprocal effects are controlled, the results still show an association between awareness of alcohol advertising and drinking knowledge, beliefs, and intentions, it is possible that different model specifications might lead to different conclusions about the exact sizes and nature of these effects. Further research should be undertaken to replicate these findings and increase confidence in them. Studies using longitudinal methods would be particularly useful.

Taken as a whole, the findings suggest that attempts to prevent or delay drinking among young people should give attention to alcohol advertising. In particular, efforts should be made to reduce the extent to which it appeals to children and to reduce their exposure to it. A number of strategies are possible to achieve these goals.

Restricting television alcohol advertising during prime viewing hours for children and limiting alcohol sponsorship of concerts and similar events that are attended by underaged youth might be considered. Addressing alcohol advertising in televised sports would be particularly important. Such actions, however, are controversial and are certain to meet considerable resistance from alcohol advertisers. Another possible approach, assuming that alcohol advertisers do not intentionally set out to influence children and adolescents, is to work cooperatively with these advertisers to reduce the extent to which alcohol advertising appeals to young people. For example, researchers could provide empirical feedback concerning the types of advertisements that are particularly appealing to youth and offer suggestions for improvements. Similar efforts with television producers and writers appear to have reduced inappropriate portrayals of drinking in prime-time fictional programs.[22]

Counteradvertising, such as that used in the California antismoking campaign,[23] also represents a reasonable strategy. However, for counteradvertising to be effective, it must compete with commercial alcohol advertisements in terms of quality, interest, and frequency of exposure. Equal time requirements or the dedication of alcohol tax funds for the production and airing of health messages may be necessary to achieve these goals. Health warning labels in alcohol advertising also may help counteract its overwhelmingly positive messages, although the evidence for the effectiveness of such labeling for children is weak.[24]

Finally, educational efforts are important in helping children develop their critical skills regarding alcohol advertisements. Many school-based prevention programs contain media components teaching young people to recognize persuasive appeals, understand the intent of advertising, and develop counterarguments to those presented in advertisements. The results of such interventions are mixed, although some appear to have had limited success when combined with other educational components.[25,26]

In the final analysis, no single preventive strategy will provide a complete answer. Ultimately, education, public policy, and community activism all will be required to ensure that children understand the role that alcohol and alcohol advertising play in our society. □

Acknowledgments

Funding for this research was provided by National Institute of Alcohol Abuse and Alcoholism grant AA06282 to the Prevention Research Center, Pacific Institute for Research and Evaluation, and by a grant from the AAA Foundation for Traffic Safety.

Preliminary versions of this paper were presented at the annual meeting of the Research Society on Alcoholism, June 1991, Marco Island, Fla, and at the annual meeting of the American Psychological Association, August 1991, San Francisco, Calif.

We thank the students, parents, and school district personnel who made this research possible. We also thank Patricia A. Madden and Diana Cassady for their assistance throughout this project.

References

1. Miller PG, Smith GT, Goldman MS. Emergence of alcohol expectancies in childhood: a possible critical period. *J Stud Alcohol.* 1990;51:343–349.
2. Goldman M, Brown SA, Christiansen BA. Expectancy theory: thinking about drinking. In: Blane HT, Leonard KE, eds. *Psychological Theories of Drinking and Alcoholism.* New York, NY: Guilford Press; 1987:181–226.
3. Christiansen BA, Goldman MS, Inn A. The development of alcohol-related expectancies in adolescents: separating pharmacological from social learning influences. *J Consult Clin Psychol.* 1982;50:336–344.
4. Christiansen BA, Goldman MS, Brown SA. The differential development of adolescent alcohol expectancies may predict adult alcoholism. *Addict Behav.* 1985;10:299–306.
5. Lipman J. Sobering view: alcohol firms put off public. *Wall St J.* August 21, 1991:B1.
6. *Surgeon General's Workshop on Drunk Driving Proceedings.* Washington, DC: US Dept of Health and Human Services; 1989.
7. *Toward a Drug-Free Generation: A Nation's Responsibility.* Washington, DC: National Commission on Drug-Free Schools; 1990.
8. Atkin C, Neuendorf K, McDermott S. The role of alcohol advertising in excessive and hazardous drinking. *J Drug Educ.* 1983;13:313–325.
9. Atkin C, Block M. *Content and Effects of Alcohol Advertising.* Washington, DC: Bureau of Tobacco, Alcohol, and Firearms; 1981. Report PB-82-123142.
10. Aitken PP, Eadie DR, Leathar DS, McNeill REJ, Scott AC. Television advertisements for alcoholic drinks do reinforce under-age drinking. *Br J Addict.* 1988;83:1399–1419.
11. Strickland DE. Advertising exposure, alcohol consumption and misuse of alcohol. In: Grant M, Plant M, Williams A, eds. *Economics and Alcohol: Consumption and Controls.* New York, NY: Gardner Press; 1983:201–222.
12. Adlaf EM, Kohn PM. Alcohol advertising, consumption and abuse: a covariance-structural modelling look at Strickland's data. *Br J Addict.* 1989;84:749–757.
13. Klitzner M, Gruenewald P, Bamberger E. Cigarette advertising and adolescent experimentation with smoking. *Br J Addict.* 1991;86:287–298.
14. McGuire WJ. Attitudes and attitude change. In: Lindzey G, Aronson E, eds. *Handbook of Social Psychology, Volume II.* New York, NY: Random House; 1985:233–346.
15. Strickland DE. Content and effects of alcohol advertising: comment on NTIS Pub. No. PB82-123142. *J Stud Alcohol.* 1984;45:87–100.
16. Madden PA, Grube JW. Alcohol advertising in sports and prime-time television. Presented at the 119th Annual Meeting of the American Public Health Association, October 1991, Atlanta, Ga.
17. Postman N, Nystrom C, Strate L, Weingartner C. *Myths, Men, and Beer: An Analysis of Beer Commercials on Broadcast Television, 1987.* Falls Church, Va: AAA Foundation for Traffic Safety; 1988.
18. Bentler PM. *EQS Structural Equations Program Manual.* Los Angeles, Calif: BMDP Statistical Software; 1989.
19. Berry WD. *Nonrecursive Causal Models.* Beverly Hills, Calif: Sage; 1984.
20. Hu L, Bentler PM, Kano Y. Can test statistics in covariance structure analysis be

trusted? *Psychological Bull.* 1992;112:351–362.
21. MacCallum RC, Roznowski M, Necowitz LB. Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bull.* 1992; 111:490–504.
22. Breed W, DeFoe J. Effecting media change: the role of cooperative consultation on alcohol topics. *J Commun.* 1982; 32:88–89.
23. Tobacco Education Oversight Committee. *Toward a Tobacco-Free California: A Master Plan to Reduce Californians' Use of Tobacco.* Sacramento, Calif: California Department of Health Services; 1991.
24. MacKinnon DP, Pentz MA, Stacy AW. The alcohol warning label and adolescents: the first year. *Amer J Public Health.* 1993; 83:585–587.
25. Botvin GJ. Substance abuse prevention research: recent developments and future directions. *J School Health.* 1986;56:369–374.
26. Botvin GJ, Tortu S. Preventing adolescent substance abuse through life skills training. In Price RH, Cowen EL, eds. *Fourteen Ounces of Prevention.* Washington, DC: American Psychological Association; 1988:98–110.

Nutritional and Lifestyle Epidemiology Course to Be Held in the Netherlands

The Department of Epidemiology and Public Health and the Department of Human Nutrition, Wageningen Agricultural University, the Netherlands, will be holding an advanced course in nutritional and lifestyle epidemiology May 30 through June 5, 1994. The following methodological issues will be dealt with on an advanced level: study design, exposure assessment, biomarkers, exposure measurement error, data analysis, bias, and effect modification. In addition, lectures will be held on such major issues as antioxidants, non-nutritive food constituents, physical activity, and passive smoking in relation to chronic diseases.

The lecturers will include, among others, Dr D. Hunter, Boston, Mass, Professor L. Kohlmeier, Chapel Hill, NC, Dr B. Margetts, Southampton, UK, Dr P. Piettinen, Helsinki, Finland, and Professor F. J. Kok and Professor W. A. van Staveren, Wageningen, the Netherlands.

Epidemiologist, nutritionists, physicians, health care professionals, and others with a basic knowledge of epidemiology are encouraged to attend.

The cost of the course is 2000 Dutch guilders, which includes hotel accommodation. For more information, contact Esther G. M. Pallast, MSc, Department of Epidemiology and Public Health, Agricultural University Wageningen, P.O. Box 238, 6700 AE Wageningen, The Netherlands; tel +31 8370 84124; fax +31 8370 82782; e-mail Esther.Pallast@medew.hegl.wau.nl.

Emotional and Behavioral Problems and Severe Academic Delays among Sheltered Homeless Children in Los Angeles County

Bonnie T. Zima, MD, MPH, Kenneth B. Wells, MD, MPH, and Howard E. Freeman, PhD

ABSTRACT

Objectives. Few studies have estimated the extent of specific emotional, behavioral, and academic problems among sheltered homeless children. The objectives of this study were to describe such problems, identify those children with the problems, and evaluate the relationship between child problems and use of physical and mental health services.

Methods. From February through May 1991, 169 school-age children and their parents living in 18 emergency homeless family shelters in Los Angeles County were interviewed. To evaluate the answers, interviewers used standardized measures of depression, behavioral problems, receptive vocabulary, and reading.

Results. The vast majority (78%) of homeless children suffered from either depression, a behavioral problem, or severe academic delay. Among children having a problem, only one third of the parents were aware of any problem, and few of those children (15%) had ever received mental health care or special education.

Conclusions. Almost all school-age sheltered homeless children in Los Angeles County have symptoms of depression, a behavioral problem, or academic delay severe enough to merit a clinical evaluation, yet few receive specific care. Programs targeted at sheltered homeless school-age children are needed to close this gap. (*Am J Public Health.* 1994;84:260–264)

Introduction

Families with children are the fastest growing homeless population, making up possibly as much as 40% of the homeless.[1,2] Almost a quarter of the homeless are children and at least 3% are unaccompanied youth.[2] On any given night, it is estimated that 68 000 to 100 000 children have no home and that an additional 186 000 are doubled up in overcrowded conditions.[3] In Los Angeles County alone, the number of homeless children is conservatively estimated to be between 32 000 and 36 000 over a 1-year period.[4]

Living in extreme poverty, homeless children experience substandard living conditions, abrupt residential transitions, and discontinuity in school.[5–7] They are also often raised by a single parent with limited education and inadequate economic and social supports,[8–10] which exacerbates the stress within the family. Earlier studies found that approximately one half of sheltered homeless children in Boston suffer from at least one developmental delay or require psychiatric evaluation for depression or anxiety,[11] and that a majority of children living in Philadelphia's public homeless shelters and New York City's hotels have extremely poor receptive vocabulary and visual motor skills.[9,12,13] Furthermore, many homeless children suffer from untreated acute and chronic physical illnesses[12,14–17] and are without a regular health care practitioner.[14,16,18]

Our study focuses on emotional disorders and severe academic delays among school-age sheltered homeless children living in one of the country's largest metropolitan counties. We targeted school-age sheltered homeless children so we could investigate those children who are most accessible to services through schools and social services while still living with their family, and we included Mexican and Central American families, a population rarely described in the homeless literature. Our objectives were to (1) describe emotional, behavioral, and academic problems among sheltered homeless children; (2) identify characteristics of homeless children with such problems; and (3) relate the use of health services and specific mental health interventions to child need.

Methods

All 22 emergency homeless family shelters that were in operation during the late winter of 1991 were identified from four Los Angeles County homeless service databases. The shelters were placed in random order and surveyed twice between February and May 1991. All families who had stayed at the shelter at least one night and had at least one child between the ages of 6 and 12 years were asked to participate. The parent who knew the child best was interviewed. If there were more than two eligible children in a family, two were randomly selected. One child with severe mental retardation

Bonnie T. Zima and Kenneth B. Wells are with the Departments of Psychiatry and Behavioral Sciences, and Howard E. Freeman was with the Department of Sociology, University of California at Los Angeles. Dr Freeman has died since the study was completed.

Requests for reprints should be sent to Bonnie T. Zima, MD, MPH, The Robert Wood Johnson Clinical Scholars Program, UCLA Medical Center, B-973 Factor Bldg, 10833 Le Conte Ave, Los Angeles, CA 90024.

This paper was accepted June 3, 1993.

Note. The views expressed are those of the authors and do not necessarily represent those of the Robert Wood Johnson Foundation.

was excluded. Interviews were conducted in Spanish or English, depending on parent and child preference; the survey instruments were translated and back-translated into Spanish. Academic skills were tested in both languages if the child was bilingual, and the best score was taken.

Measures

Sociodemographic characteristics of the parent and child were obtained from the parent. Child ethnicity was categorized as African American, Latino, non-Hispanic White, or Other. This last category included the few children identified by the parent as being biracial, Asian American, or American Indian.

Homeless history questions were adapted from the RAND Course of Homelessness Study (M. A. Burnam and P. Koegel, personal communication, March 12, 1992). Also obtained were the age the child first became homeless, the amount of time the child spent homeless in the past 12 months, the number of different places in which the child lived in the past 12 months, and the number of different times in this past year the child returned to homelessness after having a home or staying with family or friends for at least 30 continuous days. "Homeless" was defined as having no regular place to live, such as a house, an apartment, a room, or the home of a family member or friend, but instead having to stay in a shelter, an abandoned building, an all-night theater, a car, outdoors, or other such places not meant to be a permanent living space.

Both parental awareness of an emotional or learning problem and child use of health services were assessed using questions from the National Health Interview Survey, 1988 Child Health Supplement.[19] If a parent reported that the child had ever had a delay in growth or development, a learning disability, or an emotional or behavioral problem lasting 3 months or more, the parent was identified as being aware of a problem. When a problem was reported, the parent was asked whether the child had received counseling or treatment during his or her lifetime, attended special classes or a special school, or received special help at school in the past 12 months. Whether or not the child had a problem, the parent reported if the child had received routine or sickness or injury care in the past 6 months from any clinic, health center, hospital, or doctor's office.

Depression was evaluated using the Children's Depression Inventory,[20] a 27-item child self-report measure of symptoms over the past 2 weeks. Examples of symptoms are low self-esteem, poor concentration, insomnia, and a desire to commit suicide. A score above 9 is indicative of the need for psychiatric evaluation.[20] Behavioral problems were assessed using the Child Behavior Checklist.[21] This 118-item scale relies on parental report of child behavior over the past 6 months and is norm-referenced for large populations within and outside the United States. A *T* score equal to or above 60 on the total behavior scale corresponds to the 82nd percentile, the borderline clinical range warranting further evaluation. Given the residential instability and lack of structured social groups available to many homeless children, the social competence scale of the Child Behavior Checklist was not performed.

Receptive vocabulary was measured with the Peabody Picture Vocabulary Test.[22] The child was asked to point to one of four pictures that describes the spoken word best. The total score was normed for age. Three reading skills—letter-word identification, word attack, and passage comprehension—were assessed using the reading subtest of the Woodcock-Johnson Language Proficiency Battery.[23] Total reading score was normed for age and grade. Scores at or below the 10th percentile for age were considered indicative of severe delay in receptive vocabulary and reading skills. Below the 10th percentile, a child would be receiving a letter grade of F and functioning at least one grade below grade level. Both instruments had standardized Spanish versions with norms for Spanish-speaking populations.

Data Analysis

Bivariate analysis of discrete variables was performed using a chi-square test of proportions, and analysis of continuous variables was performed using analysis of variance. Linear and logistic regression analyses were used to examine whether the child's homeless history variables and demographic factors (child's age, sex, and ethnicity; parent's age, sex, marital status, education, and income) were uniquely associated with each child mental health problem, and to determine whether specific child mental health problems were uniquely associated with service use, controlling for other factors. The regression models were used to generate predicted levels of need or service use for groups of interest, such as ethnic groups, controlling for all other demographic variables. The reference population for the analysis was all 6- to 12-year-old homeless children residing in the shelters during the study period. The data were weighted by the number of eligible children per family. The significance statistics and standard errors were adjusted for the correlations in responses for children from the same family by using the intraclass correlation model.[24,25]

Results

Eighteen of the 22 (82%) of the eligible shelters agreed to participate, representing all seven geographic regions of Los Angeles County. A shelter was eligible if it allowed homeless families to stay at least one night and was an emergency shelter with a length of stay of less than 60 days. Emergency homeless family shelters ranged from missions to publicly funded facilities. Interviews were completed on 118 of 121 (98%) eligible families and on 169 (100%) eligible children.

The majority (93%) of the homeless parents living with children were women from predominantly minority backgrounds. Their mean age was 34 years; one third (34%) were married; and their average education was 11 years. The majority of families (61%) were living in extreme poverty with an annual income of less than $10 000. The average number of children per family was 3.6, with a range of 1 to 10.

One half of the children were female, and most were from minority backgrounds (Table 1). Approximately 44% of the children were African American, 35% were Latino, 14% were White, and 7% were classified as Other. Fifty-five percent of children were between the ages of 6 and 9 years, and 45% were between the ages of 10 and 12 years (mean age = 9.2 years, SD = 2.0).

The average age the child first became homeless was 7.6 years, with no significant differences between ethnic groups. More than a quarter (26%) of the children had first become homeless more than a year ago, and more than half (52%) had been homeless for more than 2 months over the past year. Almost one half (46%) had lived in more than two places over the past 12 months, and 6% had entered back into homelessness more than twice in the past year after having had a home or stayed with family or friends for at least 30 consecutive days. Compared with children from other ethnic groups, Latino children were more likely to have been homeless for more than 2 months over the past year ($P = .001$), and non-Hispanic White families were more likely

TABLE 1—Child Characteristics and Problems (%), by Ethnicity

	African American (n = 74)	Latino (n = 59)	Non-Hispanic White (n = 23)	Other (n = 12)	Total (n = 169)
Female	54	54	30	42	50
Age group					
6–9 y	51	54	60	69	55
10–12 y	49	46	40	31	45
Homelessness					
First homeless >1 year ago	28	28	32	0	26
Homeless >2 months in past year	41	71**	48	35	52
>Two residences in past year	44	39	79**	38	46
>Two returns to homelessness in past year	7	7	4	0	6
School					
Enrolled now	90	91	76	86	88
Missed >3 weeks over past 3 months	9	17	32	21	16
Problem					
Depression	34	35	48	45	37
Behavioral problem	29	28	24	28	28
Vocabulary delay	63*	42	24	21	47
Reading delay	50**	30	40	14	39

*$P \leq .05$; **$P < .01$.

TABLE 2—Child Problems (In Percentage of Children Who Received Care), by Mental Health Interventions and General Health Care

	For Mental Health		For General Health in Past 6 Months	
	Counseling/ Treatment over Lifetime	Special Classes in Past 12 Mos	Routine Care	Sickness/ Injury Care
Depression				
Present (n = 64)	23	19	72	62*
Absent (n = 105)	13	5	62	35
Behavioral problem				
Present (n = 49)	30*	17	71	54
Absent (n = 120)	11	7	64	41
Receptive vocabulary delay				
Present (n = 76)	16	14	69	43
Absent (n = 92)	17	6	62	46
Reading delay				
Present (n = 64)	19	19*	69	57*
Absent (n = 105)	15	4	64	37

*$P \leq .01$

to have lived in more than two residences over the past year ($P = .004$). Homeless history did not vary significantly by child's age or sex.

Most children (88%) were registered in school. However, 16% had missed more than 3 weeks of school over the past 3 months.

Thirty-seven percent of the children exceeded the cutoff point for depression requiring a psychiatric evaluation, and 28% scored in the borderline clinical range for a serious behavioral problem. The correlation between these scores was weak (Spearman $r = .27$). Parent and child demographics did not vary significantly between children with and without emotional and behavioral problems. Based on multiple logistic regression analyses, neither parent and child demographics nor the amount of time homeless nor the number of residences over the past 12 months predicted child depression or behavioral problems.

Forty-seven percent of homeless children scored at or below the 10th percentile in receptive vocabulary, and 39% had a severe delay in reading. That is, when asked to identify which picture corresponded to the spoken word, almost half the homeless children functioned at or below the lowest 10th percentile for children of the same age in the general population, and almost 40% demonstrated reading skills at or below the lowest decile, equivalent to an F letter grade. African-American children had significantly greater delays in receptive vocabulary ($P = .028$) and reading ($P = .001$) than homeless children from other ethnic groups, and this difference remained in the logistic regression analysis that controlled for homeless history, parental history of mental illness, and substance abuse. There were no other significant differences in parent or child demographics.

Among children requiring a psychiatric evaluation for depression, only a minority (23%) had ever received counseling or treatment in their lifetime, and even fewer (19%) had been in special education classes or received special help in school during the past 12 months (Table 2). Moreover, only 30% of those children with a behavioral problem had ever received any counseling or treatment, and even fewer (17%) had been in special education or received additional help in school in the past 12 months. Children with a behavioral problem, however, were more likely to have received counseling or treatment than those without a problem ($P = .001$). There was no significant difference in child age, sex, or ethnicity among those children receiving or not receiving any of these interventions. Only three children identified by a parent as having a problem had ever received medication for an emotional or behavioral problem, a developmental delay, or a learning disability in the past 12 months.

Likewise, only 14% of children scoring at or below the 10th percentile for age in receptive vocabulary had been in special classes or had received additional help in class during the past 12 months. Among children with a severe reading delay, only 19% had received any special education in the past year. Children with a severe reading delay, however, were more likely to receive special education than other children ($P = .001$).

Almost two-thirds (65%) of homeless children had visited a clinic, health center, hospital, or doctor's office for routine health care over the past 6 months, and fewer than half (46%) had received care

for sickness or injury within the same period. Children suffering from depressive symptoms or a severe reading delay were more likely than those without such problems to have visited a health care practitioner over the past 6 months ($P < .001$, $P = .01$).

Among the 78% of children with at least one emotional or academic problem, parents were aware of a problem for only one third of them. Moreover, of those children identified by a parent as having a problem, only one half had received any type of mental health or academic intervention.

Discussion

Most sheltered homeless children in Los Angeles County require clinical evaluation for serious emotional and behavioral problems and severe academic delays. The high prevalence of problems is alarming but not surprising, given that many homeless children not only possess multiple traditional risk factors for mental disorders, such as poverty, stress, and single-parent families,[26–33] but also experience stressors perhaps unique to being homeless, such as frequent moves and discontinuity in school at a relatively young age.

The sheltered homeless children studied here were almost 20 times more likely to have depressive symptoms than prepubescent children in the general population.[34] Clinical assessment of symptomatic homeless children is needed to determine whether their depressive symptoms correspond to a major mental disorder or are a response to the stress of being homeless, a transient phenomenon, or an artifact of the screening instrument. In an extensive literature review, we found few clues on the validity of the depression measure in low-income minority children.

Similarly, sheltered homeless children in our study were 1.5 times more likely to have symptoms of a behavioral disorder than those children in the measure's large normative general population. As noted elsewhere, race and socioeconomic status have little effect on scale scores of the Child Behavior Checklist,[21] so our finding is unlikely to be an artifact of the screening instrument.

Homeless children were also approximately four times more likely to score at or below the 10th percentile in receptive vocabulary and reading than children of similar age in the general population. That is, a substantial number of homeless children, although attending school, did not possess average grade-level reading and receptive vocabulary skills to allow them to learn competitively alongside their housed counterparts. Greater academic delays found among African-American homeless children could reflect a cultural bias in the testing instruments[35]; however, even when tested and scored in their preferred language, a high percentage of Latino children, as well as of non-Hispanic Whites, had serious academic delays. Both academic measures had large normative samples for English- and Spanish-speaking children.[22,23]

Yet despite this high need, we found that few homeless children receive treatment. Moreover, a target population of those children most likely to require services could not be identified; demographic characteristics and homeless history parameters did not predict child emotional disorders or academic delays, probably because of the high prevalence of problems and important predictor variables such as poverty status. Thus, service provision and not health screening appears to be the priority for school-age sheltered homeless children. Our study findings offer some insights into how such provision may be made.

Residential instability, a paucity of community mental health clinics, erratic school attendance, and the lack of parental awareness of a child's problem may make it particularly difficult for homeless children to receive services. Hence, mental health and academic interventions for homeless children should be adapted to accommodate the crisis nature of sheltered homeless families and include parental education and mechanisms for continued care once the family obtains stable housing. Furthermore, both the relatively high level of contact with a health care practitioner for routine general health care and the greater likelihood that depressed children receive care for sickness suggest that the general medical sector could be a potentially important point of contact for serving homeless children.

This study has two main limitations. One is the absence of a comparison group of low-income, housed, and predominantly minority children. However, a suitable control group, one that would have a similar distribution of risk and protective factors except for being homeless, could not be identified because of the unique and varying combinations of social and economic disadvantages among homeless families. The other limitation—the exclusion of nonsheltered homeless children—may have resulted in the prevalence of child problems being underestimated since the number of homeless families unable to obtain emergency shelter is estimated to be greater than the number of sheltered homeless families.[3] Furthermore, the study's cross-sectional design and age restriction prohibit conclusions about risk factors for behavioral problems and developmental delays among the larger homeless child population under 6 years of age.[36]

Despite these limitations, the prevalence of serious emotional problems and severe academic delays reported here suggests that programs for homeless children are needed, regardless of the level of need in any comparison group. The strength of our findings is supported by comparable results from earlier studies based in other major cities[9,11,12] and by a high participation rate from a diversity of emergency homeless shelters in a heavily populated metropolitan county. Mental health and educational services for homeless children should be a component of a comprehensive package of services that assist families in obtaining permanent housing. These programs could require participation of the schools, shelters, and routine health care practitioners, as well as inclusion of parental education. Without such timely intervention, however, the emotional and behavioral problems and severe academic delays of homeless children will go untreated, threatening the development of these children at a critical period in their lives and robbing them of the fundamental opportunity to gain the emotional tools and academic skills necessary for fully productive and independent lives. □

Acknowledgments

This research was conducted under the Robert Wood Johnson Clinical Scholars Program at the University of California at Los Angeles.

The authors would like to express their gratitude to the Robert Wood Johnson Foundation for its financial assistance; to Drs Robert H. Brook, M. Audrey Burnam, Arlene Fink, Steven Forness, Ralph Frerich, Milton Greenblatt, and Paul Koegel for their assistance in research design; to Barbara Genovese, Enrique Godinez, Marlene Nishimoto-Horowitz, Adam Oberweiser, Monica Paez, and Ruth Stearn for their diligent work collecting data; to Dr Barbara Leake at the University of California at Los Angeles and Dr Naihua Duan and Bernadette Benjamin at the RAND Corporation in Santa Monica, Calif, for statistical consultation; and to shelter staff and residents for their strong support and cooperation that made this study possible.

This paper is dedicated to the memory of Dr Howard E. Freeman, an outstanding mentor and teacher.

References

1. *A Status Report on Hunger and Homelessness in America's Cities: 1993—A 26 City Survey.* Washington, DC: US Conference of Mayors; 1993.
2. *A Status Report of Homeless Families in America's Cities: A 29 City Survey.* Washington, DC: US Conference of Mayors; 1987.
3. *Children and Youths—About 68 000 Homeless and 186 000 in Shared Housing at Any Given Time.* Washington, DC: US General Accounting Office; 1989.
4. *The Number of Homeless People in Los Angeles County—July 1989–June 1990.* Los Angeles, Calif: Shelter Partnership; 1990.
5. *Shut Out: Denial of Education to Homeless Children.* Washington, DC: National Law Center on Homelessness and Poverty; 1990.
6. Rafferty Y, Shinn M. The impact of homelessness on children. *Am Psychol.* 1991;46: 1170–1179.
7. Rafferty Y, Rollins N. *Learning in Limbo: The Educational Deprivation of Homeless Children.* New York, NY: Advocates for Children; 1989. ERIC document reproduction ED 312 363.
8. Bassuk EL, Rubin L, Lauriat AS. Characteristics of sheltered homeless families. *Am J Public Health.* 1986;76:1097–1101.
9. Fox SJ, Barrnett RJ, Davies M, Bird HR. Psychopathology and developmental delay in homeless children: a pilot study. *J Am Acad Child Adolesc Psychiatry.* 1990;29: 732–735.
10. Wood D, Valdez RB, Hayashi T, Shen A. Homeless and housed families in Los Angeles: a study comparing demographic, economic, and family function characteristics. *Am J Public Health.* 1990;80:1049–1052.
11. Bassuk EL, Rosenberg L. Psychosocial characteristics of homeless children and children with homes. *Pediatrics.* 1990;85: 257–261.
12. Parker RM, Rescorla LA, Finkelstein JA, Barnes N, Holmes JH, Stolley PD. A survey of the health of homeless children in Philadelphia shelters. *Am J Dis Child.* 1991;145:520–526.
13. Rescorla L, Parker R, Stolley P. Ability, achievement, and adjustment in homeless children. *Am J Orthopsychiatry.* 1991;61: 210–220.
14. Miller DS, Lin EHB. Children in sheltered homeless families: reported health status and use of health services. *Pediatrics.* 1988;81:668–673.
15. Alperstein G, Rappaport C, Flanigan JM. Health problems of homeless children in New York City. *Am J Public Health.* 1988; 78:1232–1233.
16. Bass JL, Brennan P, Mehta KA, Kodzis S. Pediatric problems in a suburban shelter for homeless families. *Pediatrics.* 1990;85:33–38.
17. American Academy of Pediatrics Committee on Community Health Services. Health needs of homeless children. *Pediatrics.* 1988;82:938–940.
18. Hu DJ, Covell RM, Morgan J, Arcia J. Health care needs for children of the recently homeless. *J Community Health.* 1989;14:1–8.
19. *National Health Interview Survey: Child Health Supplement, 1988* (Computer data files). 1st ed. Washington, DC: National Center for Health Statistics; 1988. 1 Data File and Code Book.
20. Kovacs M. The children's depression inventory (CDI). *Psychopharmacol Bull.* 1985;21(special ed):995–998.
21. Achenbach TM, Edelbrock CS. *Manual for the Child Behavior Checklist and Revised Behavior Profile.* Burlington, Vt: University of Vermont Department of Psychiatry; 1983.
22. Dunn LM, Dunn LM. *Peabody Picture Vocabulary Test—Revised.* Circle Pines, Minn: American Guidance Service; 1981.
23. Woodcock RW. *Woodcock-Johnson Language Proficiency Battery.* Allen, Tex: DLM Teaching Resources; 1984.
24. Searle SR. *Linear Models.* New York, NY: John Wiley and Sons; 1971.
25. Huber PJ. *The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions.* Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, Calif: University of California Press; 1967;1:221–233.
26. Offord DR, Boyle MH, Racine YA, et al. Outcome, prognosis, and risk in a longitudinal follow-up study. *J Am Acad Child Adolesc Psychiatry.* 1992;31:916–923.
27. Velez CN, Johnson J, Cohen F. A longitudinal analysis of selected risk factors for child psychopathology. *J Am Acad Child Adolesc Psychiatry.* 1989;28:861–864.
28. Bird HR, Gould MS, Yager T, Staghezza B, Canino G. Risk factors for maladjustment in Puerto Rican children. *J Am Acad Child Adolesc Psychiatry.* 1989;28:847–850.
29. Goodyer IM, Kolvin I, Gatzanis S. The impact of recent undesirable life events on psychiatric disorders in childhood and adolescence. *Br J Psychiatry.* 1987;151:179–184.
30. Stanger C, McConaughy SH, Achenbach TM. Three year course of behavioral/emotional problems in a national sample of 4–16 year olds: II. predictors of syndromes. *J Am Acad Child Adolesc Psychiatry.* 1992;31:941–950.
31. Jensen PS, Richters J, Ussery T, Bloedau L, Davis H. Child psychopathology and environmental influences: discrete life events versus ongoing adversity. *J Am Acad Child Adolesc Psychiatry.* 1991;30:303–309.
32. Moilanen I, Rantakallio P. The single parent family and the child's mental health. *Soc Sci Med.* 1988;27:181–186.
33. McLoyd VC. The impact of economic hardship on Black families and children: psychologic distress, parenting, and socioemotional development. *Child Devel.* 1990;61:311–346.
34. Kashani JH, Sherman DD. Childhood depression: epidemiology, etiological models, and treatment implications. *Int Psychiatry Clin.* 1988;6:1–8.
35. Bracken BA, Prass DP. Comparison of the PPVT, PPVT-R and intelligence tests used for placement of Black, White, and Hispanic EMR students. *J Sch Psychol.* 1982; 19:304–311.
36. *The State of America's Children 1991.* Washington, DC: Children's Defense Fund; 1991.

Risk Factors for Homelessness among Schizophrenic Men: A Case–Control Study

ABSTRACT

Objectives. To identify risk factors for homelessness among the severely mentally ill, we conducted a case–control study of 100 indigent schizophrenic men meeting criteria for literal homelessness and 100 such men with no homeless history.

Methods. Subjects were recruited from shelter, clinic, and inpatient psychiatric programs in Upper Manhattan. Clinical interviewers employed standardized research instruments to probe three domains of risk factors: severity of mental illness, family background, and prior mental health service use.

Results. Homeless subjects showed significantly higher levels of positive symptoms, higher rates of a concurrent diagnosis of drug abuse, and higher rates of antisocial personality disorder. Homeless subjects experienced greater disorganization in family settings from birth to 18 years and less adequate current family support. Fewer homeless subjects than subjects in the never-homeless comparison group had a long-term therapist. These differences remained when demographic variables were adjusted statistically.

Conclusions. Homeless schizophrenic men differed from their domiciled counterparts in all three domains we investigated; family background, nature of illness, and service use history. Findings are discussed in relation to policy and programs for the severely mentally ill. (*Am J Public Health.* 1994;84:265–270)

Carol L. M. Caton, PhD, Patrick E. Shrout, PhD, Paula F. Eagle, MD, Lewis A. Opler, MD, PhD, Alan Felix, MD, and Boanerges Dominguez, MS

Introduction

A recent governmental report estimates that 1 in 20 of the severely mentally ill experience homelessness.[1] Using different definitions of homelessness or housing instability, empirical studies of state mental hospital constituents have found that from 9% to 28% have been homeless.[2–4] The crisis nature of homelessness, both for the individual and for society, demands a clearer understanding of why some of the severely mentally ill become homeless while others do not.

Previous studies in which differences between the homeless and the residentially stable have been explored suggest that the homeless have more psychiatric disorders meeting standard diagnostic criteria[5,6] and higher rates of hospitalization and arrest.[5,7] Studies focused on defined psychiatric populations have found that the homeless are more likely to abuse alcohol[3,8] and/or drugs,[3,4,6] have higher symptom levels,[3,7,8] and are less likely to comply with prescribed treatments.[3,8] The use of different definitions of homelessness, the employment of varied approaches to the assessment of psychiatric disorder, and the selection of different subgroups of the homeless population in these investigations limit their generalizability.

Attempts to identify areas of differences between the homeless and the never homeless outside of the illness domain are more limited. Studies of childhood antecedents of adult homelessness have revealed that homelessness (or its duration) is associated with placement in foster care[9,10] or group homes,[10] physical abuse,[9] and runaway episodes.[10] Although the breakdown of family ties has been thought to play a role in the genesis of homelessness, perceptions of the kinship bond in adulthood have not been found to be as important as patient characteristics in distinguishing the homeless from the nonhomeless.

In spite of the discourse on deinstitutionalization of mental health services as a possible factor in homelessness among the severely mentally ill,[11,12] the service use histories of homeless and never-homeless persons have not been compared with regard to access to and utilization of community-based treatment and support services. Although the literature suggests that multiple factors may separate the homeless from the never homeless, there has been no prior attempt to incorporate multiple risk factors into a single study.

We recently completed a case–control study of indigent schizophrenic men in New York City designed to test hypotheses about differences between the homeless and never-homeless mentally ill. To facilitate an adequate test of hypotheses about differences between the homeless and never homeless, we elected to focus our study on a homogeneous group of indigent adults suffering from schizophrenia. Schizophrenia is the most common diagnosis leading to chronic mental illness[13] and is widely prevalent among those living in shelters.[1,14,15] Because four of five homeless adults are male,[16] it is

Carol L. M. Caton, Paula F. Eagle, Lewis A. Opler, Alan Felix, and Boanerges Dominguez are with the Department of Psychiatry and School of Public Health, College of Physicians and Surgeons, Columbia University, New York, NY. Patrick E. Shrout is with the Department of Psychology, New York University.

Requests for reprints should be sent to Carol L. M. Caton, PhD, Department of Psychiatry, Columbia University, 722 W 168th St, Box 114, New York, NY 10032.

This paper was accepted August 23, 1993.

possible that pathways to homelessness are gender-specific. Therefore, our sample was stratified by gender. Findings from our study of women will be reported later.

We probed three domains of influence on homelessness: severity of illness, family background, and prior mental health service use. We hypothesized that schizophrenic men who became homeless would have poorer pre-illness social functioning, higher levels of the positive and negative symptoms of schizophrenia, greater abuse of drugs and alcohol, and more antisocial personality disorder. In addition, we hypothesized that the homeless would have experienced greater family disorganization in childhood, would have imposed a greater caretaker burden after becoming mentally ill, and would have less current family support. Finally, we hypothesized that the homeless would be less adherent to prescribed treatment and would have less help from the mental health system in finding adequate housing and procuring long-term follow-up care.

Methods

We employed a case–control design with 100 men in the (sheltered) homeless group and 100 men in the never-homeless group. All 200 subjects had experienced at least one psychiatric hospitalization and were between the ages of 18 and 44 years. To be eligible for inclusion, the men were required to meet DSM-III-R criteria for schizophrenia or schizoaffective disorder, determined through a research diagnostic interview, the Structured Clinical Interview for DSM-III-R (SCID).[17] Voluntary informed consent was elicited from each subject after a treating clinician attested to the subject's ability to fully comprehend the implications of participation in our research.

Case subjects met criteria for literal homelessness, meaning that they had no fixed abode and were forced to sleep in the street or in shelters.[18] They were recruited by a mental health day treatment program located on-site at a municipal shelter for homeless men in Upper Manhattan.

Control subjects had no lifetime history of literal homelessness, as determined by a screening instrument.[4] They were discharge-ready inpatients or outpatients recruited from Columbia University–affiliated psychiatric services in Upper Manhattan. More than four of five control subjects resided in a family setting, usually with a parent or sibling.

Because Blacks have been found to be overrepresented among the sheltered homeless in New York City,[19] we made every effort to match case and control subjects on ethnicity (Black, White, Hispanic). When we encountered difficulty in finding an adequate number of Black subjects who had never been homeless, we abandoned our matching strategy. Reasons for the paucity of Black never-homeless subjects constitute a topic in need of further study.

A total of 248 men were asked to participate in the study. Eleven (4.4%) refused, and 22 (8.9%) dropped out before completing the interview battery. Four of five dropouts were homeless men who precipitously left the shelter without letting the mental health staff know how to reach them. Fifteen (6.1%) of the homeless subjects with completed interviews were eliminated from the study because of inconsistent or poor-quality data that could not be improved with a reinterview. In such cases the subject was either too delusional or too disorganized to function as a reliable informant. It is possible that the more disturbed subjects were not included in the final sample.

Subjects were interviewed by mental health clinicians specially trained to administer the assessment battery. Clinical and anamnestic data were verified with information contained in the clinical case record. A family member was also interviewed for 57% of the homeless and 68% of the never homeless. The most common reason for the lack of a family interview was the subject's estrangement from kin.

Instruments

The following research instruments were used to assess key study variables:

Pre-illness social functioning was rated with the Los Angeles Social Attainment Scale, a seven-item instrument focused on peer relationships and social participation in late adolescence.[20] The alpha reliability coefficient for this scale is .80.

The positive (active psychotic symptoms such as delusions, hallucinatory behavior, and conceptual disorganization) and negative (deficit symptoms such as blunted affect, emotional withdrawal, and difficulty with abstract thinking) dimensions of schizophrenia were assessed with the Positive and Negative Syndrome Scale, a 30-item rating scale evaluating symptoms present during the 7-day period prior to the interview.[21] The alpha reliability coefficient was .79 for the positive scale, .84 for the negative scale, and .83 for the general psychopathology scale.

Current and lifetime alcohol and drug abuse or dependence were evaluated with the SCID, which yields information on the extent to which heavy use of alcohol or seven classes of drugs meets criteria for a diagnosis based on standards commonly applied in psychiatry.[17]

Antisocial personality disorder was evaluated with the SCID-II, the segment of the SCID dealing with personality disorders.[17] This instrument probes the presence of conduct problems—such as running away from home, being truant from school, initiating physical fights, or belonging to a gang—occurring before the age of 15, as well as patterns of irresponsible, destructive, or illegal activity carrying over into adulthood. For a subject to meet criteria for a diagnosis of antisocial personality disorder, a conduct disorder must have been present in early adolescence.

Family disorganization in childhood was evaluated with 4-point rating scales contained in the Community Care Schedule.[22] These ratings are based on carefully defined anchor points for the assessment of nurturing constancy, residential stability, adequacy of income, dependence on public assistance, family violence, parental criminality, parental mental illness, and parental substance abuse—the components of an index of family disorganization. Information for this instrument was elicited from both the subject and a family member (whenever possible) to obtain the most complete picture of family life during the subject's childhood. The alpha reliability coefficient for the seven-item index of family disorganization is .69.

Current family support was rated on a scale of adequacy, based on the degree of support and assistance available from family members regarding money, shelter, food, clothing, advice, and companionship.[22] Information for this rating was obtained from the subject and corroborated by a relative when possible.

Caretaker burden was assessed with a semistructured interview that yielded a 3-point rating of overall burden.[23] Unlike other family variables, family burden required an assessment from a family member with firsthand knowledge of the family experience of living with the patient. Family burden data were missing for one third of never-homeless and nearly one half of homeless case subjects.

Prior service use was explored with items contained in the Community Care Schedule[22] that were designed to elicit detailed information on medication adherence patterns, long-term follow-up care, and housing placement at hospital discharge. Medication adherence was rated on a 4-point scale on the basis of the sub-

ject's self-report. Long-term follow-up care was defined in terms of the number of months in outpatient treatment with the same therapist.

Statistical Methods of Analysis

The case subjects (homeless) were compared with the control subjects (never homeless) with respect to three classes of key variables. In the illness domain, there were six variables: premorbid social attainment scale, positive symptom score, negative symptom score, and binary indicators of alcohol abuse, drug abuse and antisocial personality disorder. In the family domain, there were three variables: family disorganization index, adequacy of family support, and assessment of burden. In the service use domain, there were two variables: medication adherence and existence of a long-term therapist.

First we compared the case and control subjects without taking into account sampling characteristics of the groups, and then we compared them after making statistical adjustments for demographic variables that could possibly account for observed differences. Statistical adjustments were made by means of logistic regression, with the binary case/control variable treated as the outcome and the risk variables and possible confounders treated as explanatory variables. Possible confounders were chosen from those demographic variables shown in Table 1 that were at all related to the case/control variable (at the $P < .15$ level). These included place of birth, New York City residence history, ethnicity, and veteran status. We did not adjust for current income variables, because they can be expected to be related to homelessness structurally rather than incidentally.

To facilitate the comparison of adjusted and unadjusted associations between the risk variables and the case/control distinction, we presented both using results from logistic regression analyses. The test statistic we present is the likelihood ratio chi-square from the logistic analyses.[24] For the unadjusted case/control comparisons the pattern of significance results from the likelihood ratio chi-square is the same as would be obtained by means of t tests or Pearson chi-square statistics.

Results

Demographic Profile

The homeless (case) group and never-homeless (control) group had many traits in common, including median age (32 years); marital status (four fifths had never married or were currently single); median level of education (11 years); and employment status (more than four fifths were unemployed). Moreover, most (89% of the homeless and 76% of the never homeless) were members of ethnic minorities. However, there were more Blacks among the homeless (66% vs 22%) and more Hispanics among the never homeless (54% vs 23%). Ethnic differences were reflected in differences in religious background (Catholics predominated among the never homeless, whereas Catholics and Protestants were nearly equally represented among the homeless) and place of birth. Nearly twice as many never homeless as homeless (39% vs 21%) were foreign-born.

Both groups were heavily dependent on public assistance. Nearly four fifths of the never homeless and three fifths of the homeless were receiving income from entitlements, averaging about $430 per month. Thirteen percent of the never homeless and 19% of the homeless earned an average of $100 per month at unskilled jobs. Thirty-three percent of the never homeless and 8% of the homeless also received some financial support from their families. For subjects who had ever worked, the longest duration of steady employment was an average of 12 months. Unemployed subjects had been out of work for an average of 4 to 5 years. Nine percent of the never homeless and 22% of the homeless had served in the armed forces.

Family History and Illness Onset

Thirty-five percent of the never homeless and 23% of the homeless had lived with both biological parents from birth to 18 years of age. The majority in both groups experienced one or more changes in the primary nurturing adult in childhood and adolescence. One fourth of the never homeless and more than one third of the homeless had experienced two or more such changes. One never-homeless subject and six homeless subjects experienced foster care placement; four homeless subjects had been placed in a group home. Four never-homeless subjects and 20 homeless subjects had a father, mother, or sibling with a history of homelessness.

The median age of onset of psychiatric disorder was 21 years for the never homeless and 20 years for the homeless. Never-homeless subjects experienced their first psychiatric hospitalization at a median age of 22 years; for homeless subjects it was 21 years. Most subjects had extensive hospitalization histories. Fifty-eight percent of the never homeless and 56% of the homeless had been hospitalized in the 12-month period preceding the

TABLE 1—Background Characteristics of Never-Homeless and Homeless Schizophrenic Men

	Never Homeless (n = 100)	Homeless (n = 100)
Current age, y		
Mean	33.86	32.95
Median	32.00	32.00
Mode	32.00	30.00
SD	9.15	6.39
Education, y		
Mean	11.01	11.02
Median	11.00	11.00
Mode	12.00	11.00
SD	2.99	2.25
Place of birth, %		
United States	61	79
Foreign	39	21
Residence history (residence for the greatest percentage of time from birth to 18 years), %		
New York City area	67	66
Other	33	34
Race/ethnicity, %		
Black	22	66
Hispanic	54	23
White	24	11
Religion, %		
Protestant	20	43
Catholic	61	40
Other	11	11
None	8	6
Marital status, %		
Single	78	79
Married/conjugal	7	3
Separated/divorced	14	16
Widowed	1	1
Veteran status, %		
Yes	7	22
No	93	78
Employment status, %		
Employed	10	11
Unemployed	90	89
Income from entitlements, %		
Yes	78	59
No	22	41
Earned income, %		
Yes	13	19
No	87	81
Income from family, %		
Yes	33	8
No	67	92

TABLE 2—Test of Key Hypotheses about Differences between Homeless and Never-Homeless Urban Schizophrenic Men

	Never Homeless (n = 100)		Homeless (n = 100)		Unadjusted[a]		Adjusted[b]	
	Mean	SD	Mean	SD	LRT	P	LRT	P
Illness domain								
Social attainment score	19.24	6.93	19.92	5.97	0.54	.46	0.03	.87
Positive symptoms score	13.00	5.82	15.73	6.72	9.28	<.01	4.57	<.05
Negative symptoms score	15.53	6.04	15.99	6.29	0.28	.60	0.56	.45
Alcohol abuse (0 = no, 1 = yes)	0.44	0.50	0.49	0.50	.50	.48	0.16	.69
Drug abuse (0 = no, 1 = yes)	0.44	0.50	0.77	0.42	23.33	<.001	17.55	.001
Antisocial personality disorder (0 = no, 1 = yes)	0.09	0.29	0.42	0.50	30.54	<.001	17.99	<.001
Family domain								
Index of family disorganization	11.88	3.29	13.98	4.65	12.18	<.001	4.64	<.05
Adequacy of family support	1.42	0.88	2.28	1.20	65.81	<.001	36.6	.001
Overall assessment of burden	2.03	0.63 (67 cases)	1.95	0.65 (55 cases)	0.54	.46	0.25	.62
Service use domain								
Medication adherence	1.52	0.83	1.96	1.09	10.24	<.01	3.13	<.10
Long-term therapist	36.78	36.97	19.36	23.08	14.77	<.01	7.86	<.01

[a]Likelihood ratio test (LTR) and P value from logistic regression models with no other variables held constant.
[b]LRT and P value from logistic regression models holding constant immigration status, New York City residence history, ethnicity, and veteran status.

interview. The median number of hospital days in this time period was 60 for the never homeless and 43 for the homeless subjects. At the termination of the most recent hospitalization episode, 11 homeless subjects and 1 never-homeless subject had been discharged against medical advice. Ninety-six percent of the never homeless and 83% of the homeless had been given a recent prescription for neuroleptic medication (the class of psychotropic drugs often used to treat schizophrenia).

Subjects with substance abuse problems had initiated alcohol and drug use in midadolescence. The median age of onset of drug abuse (any drugs) was 18 years for never-homeless subjects (44%) and 16 years for homeless subjects (77%). The median age of onset for alcohol abuse was 18 years for both groups (44% of the never homeless and 49% of the homeless). Fifteen percent of the never homeless and 34% of the homeless had received some form of substance abuse treatment in the 12-month period preceding the interview.

Thirteen percent of the never homeless and 46% of the homeless had had severe conduct problems before the age of 15 years. Six percent of the never homeless and 28% of the homeless reported that they had run away from home at least once during early adolescence. Nine percent of the never homeless and 42% of the homeless were given a research diagnosis of adult antisocial personality disorder, and 25% of the never homeless and 72% of the homeless had jail or prison histories.

Distinctions between the Homeless and the Never Homeless

Table 2 summarizes tests of study hypotheses about the differences between never-homeless and homeless schizophrenic men. The tests are presented in two forms, one that is unadjusted for possible case–control differences that may confound the group comparisons and one that adjusts for demographic factors that may be related to selection into the study.

Within the illness domain, there were no significant differences between the homeless and the never homeless on the Los Angeles Social Attainment Scale scores in either unadjusted or adjusted tests. Average ratings on peer socialization and social participation in late adolescence were at the "poor premorbid" level in both groups. Homeless subjects did show significantly higher levels of positive symptoms than did the never-homeless subjects ($P < .01$, unadjusted; $P < .05$ adjusted). However, there were no major differences in negative symptom levels. There were no meaningful differences in the proportion of subjects in each group with a concurrent diagnosis of alcohol abuse, but a significantly larger number of homeless subjects had a concurrent diagnosis of drug abuse ($P < .01$ for both unadjusted and adjusted tests). Similarly, a significantly larger number of homeless subjects had a concurrent diagnosis of antisocial personality disorder ($P < .01$ for both unadjusted and adjusted tests).

Within the family domain, the mean score on the index of family disorganization was higher, indicating greater impairment, for homeless subjects ($P < .01$, unadjusted; $P < .05$, adjusted). Family support was less adequate for the homeless ($P < .01$ for both tests). However, there were no significant differences in the overall levels of caretaker burden.

In terms of service use issues, the homeless initially appeared to adhere less to prescribed medication regimens ($P < .01$, unadjusted), but that difference did not withstand adjustment for potential confounders ($P < .10$). Fewer homeless subjects had a long-term therapist, an indicator of continuity of care ($P < .01$, both adjusted and unadjusted).

At discharge from the most recent psychiatric hospitalization, 50% of the homeless (undomiciled before hospitalization) were discharged to a shelter or the streets. The 95% confidence interval for the proportion P is between 40% and 60%.

Table 2 shows how each of the individual risk variables relates to homelessness after demographic factors that might be related to selection are controlled. We carried out additional analyses to determine the extent to which these variables accounted for the same variation in the homelessness distinction. When all the risk variables in Table 2 were entered simultaneously in a logistic model fitting log odds of homelessness (along with the demographic control variables), one variable in each domain remained either significant or a trend. Still significant by the Wald statistic were drug abuse (from the illness domain) (Wald = 5.28, $df = 1$, $P < .05$)

and adequacy of family support (from the family domain) (Wald = 14.41, $df = 1$, $P < .001$), whereas having a long-term therapist (from the service use domain) was reduced to a trend (Wald = 3.07, $df = 1$, $P < .10$). These three variables seem to reflect different aspects of the differences between the homeless and never-homeless groups.

Two variables under the illness domain that are significant in Table 2 are not significant when adjusted for the other risk variables. Positive symptoms is no longer a discriminating variable when other illness domain variables are controlled, but antisocial personality disorder loses significance only when adequacy of current family support is adjusted.

Adequacy of family support also appears to be the variable that accounts for the effects of family disorganization index and having a long-term therapist. These variables are significant in Table 2, but they do not appear to differentiate the homeless from the never homeless when adequacy of family support is adjusted.

Discussion

In this study we have attempted to identify factors that differentiate economically disadvantaged schizophrenics who live in private homes from those who are homeless and who live in public shelters. Although not definitive in terms of causal pathways, a case–control design such as ours is an efficient way to spot potential risk factors for homelessness. Some variables we explored preceded the first homeless experience and may play a causal role. In addition to childhood antecedents, our data indicate that drug abuse and antisocial personality disorder had their onset prior to homelessness. The cross-sectional nature of our study prevents us from determining whether other variables are a cause or a consequence of homelessness. Our data show an association between family disorganization in childhood and poor family support in adulthood in assessing the risk of homelessness. However, a longitudinal approach is needed to establish that breakdown of family relationships plays a causal role in homelessness.[7] Higher current levels of positive symptoms and the lack of a long-term therapist could be either a cause or a consequence of homelessness. Higher symptom levels can precipitate behaviors that undermine housing stability, but the vicissitudes of street or shelter living can increase distress and disrupt usual patterns of mental health care. Our data on hospital discharge practices for homeless subjects suggest that the pattern of discharging patients to shelters or streets functions to perpetuate homelessness.

The sampling design of our study formally limits us to generalizing our findings only to persons who have the same demographic and clinical characteristics as our study subjects and who are enrolled in mental health treatment programs similar to those from which we drew our sample. However, the pattern of findings suggests that they may be more general, particularly among the severely mentally ill. We found that the homeless schizophrenic men differed from their domiciled counterparts in all three domains we investigated; family background, nature of illness, and service use history.

Schizophrenic men who came from families with high levels of family disorganization, characterized by lack of nurturing constancy, residential instability, economic inadequacy, dependence on public assistance, family violence, mental illness, substance abuse, or criminality in a parent, were more likely to be in the homeless group than in the never-homeless group, even when demographic variables such as place of birth and ethnicity were controlled. Our findings broaden the scope of childhood antecedents of adult homelessness to include family economic problems and parental pathology, the latter being strongly associated with out-of-home placement.[25] However, the logistic regression analysis revealed that poor family support is a more important risk factor for homelessness than childhood antecedents. We thought it reasonable to ask whether the overall burden experienced by the family in taking care of the schizophrenic relative might differentiate the homeless from the never homeless. Our negative findings should be interpreted cautiously because our ability to measure family burden was limited.

The men who ended up in our homeless group do appear to have a strikingly different profile of psychopathology. They were more likely to have positive symptoms of psychosis and to have concurrent drug abuse and antisocial personality disorder. The logistic regression analysis revealed that drug abuse was the most important risk factor in the illness domain. There were no significant differences between the homeless and the never homeless in premorbid social adjustment, negative symptoms, or alcohol abuse. Thus, the homeless in our study were not more severely ill in all dimensions of psychopathology.

The finding that 50% of the homeless (undomiciled before hospitalization) were discharged to a shelter or the streets indicates that the homeless in our study had less access to adequate care. Although this finding may reflect local mental health practices, similar findings have emerged from studies carried out in Chicago[26] and rural Ohio.[27] The logistic regression analysis revealed that the most important risk factor in the service use domain was not having a long-term therapist. Future work will have to determine whether lack of a long-term therapist is an access problem or a consequence of resistance to treatment.

Clearer identification of risk factors for homelessness can lead to the development of preventive intervention strategies. For example, thorough assessment of patterns of substance use, family support, and prior outpatient service use in concert with routine discharge planning can identify those who require additional services and supportive housing. Patients with co-disorders can be singled out for intensive case management to improve the successful coordination of mental health and substance abuse treatment services.[28–31] Finally, although risk factors aid in identifying those individuals most vulnerable to homelessness, they do not obviate the need for policymakers at all levels of government to address the chronic shortage of affordable housing for indigent Americans. □

Acknowledgments

The research on which this report is based was supported by the National Institute of Mental Health, grant MH44705.

Our thanks to our dedicated and talented research support staff, including Luz Romero, BS (who also assisted in the preparation of the manuscript); Andrea Cassells, MPH; Louis Caraballo, BA; Diane Engel, CSW; Lee Futrovsky, PhD; Barbara Holton, CSW; Elizabeth Margoshes, PhD; and Mary Ellen Russell, MS. We are also grateful to Richard Jed Wyatt, MD, and Jerome K. Myers, PhD, for their critical reading of an earlier draft.

References

1. Federal Task Force on Homelessness and Severe Mental Illness. *Outcasts on Main Street*. Washington, DC: Interagency Council on the Homeless; 1992:18.
2. Mowbray CT, Johnson US, Solar A. Homelessness in a state hospital population. *Hosp Community Psychiatry*. 1987; 38:880–882.
3. Drake RE, Wallach MA, Hoffman JS. Housing instability and homelessness among aftercare patients of an urban state hospital. *Hosp Community Psychiatry*. 1989;40:46–51.
4. Susser ES, Lin SP, Conover SA. Risk fac-

tors for homelessness among patients admitted to a state mental hospital. *Am J Psychiatry*. 1991;148:1659–1664.
5. Fischer PJ, Shapiro S, Breakey WR, Anthony JC, Kramer M. Mental health and social characteristics of the homeless: a survey of mission users. *Am J Public Health*. 1986;76:519–524.
6. Koegel P, Burnam A, Farr RK. The prevalence of specific psychiatric disorders among homeless individuals in the inner city of Los Angeles. *Arch Gen Psychiatry*. 1988;45:1085–1092.
7. Tessler RC, Gamache GM, Rossi PR, Lehman AF, Goldman HH. The kindred bonds of mentally ill homeless persons. *N Engl J Public Policy*. 1992;8:265–280.
8. Drake RE, Wallach MA, Teague GB, Freeman DH, Paskus TS, Clark TA. Housing instability and homelessness among rural schizophrenic patients. *Am J Psychiatry*. 1991;148:330–336.
9. Winkleby MA, Rockhill B, Jatulis D, Fortmann SP. The medical origins of homelessness. *Am J Public Health*. 1992;82:1394–1398.
10. Susser ES, Lin SP, Conover SA, Struening EL. Childhood antecedents of homelessness in psychiatric patients. *Am J Psychiatry*. 1991;148:1026–1030.
11. Lamb HR. Deinstitutionalization and the homeless mentally ill. *Hosp Community Psychiatry*. 1984;35:899–907.
12. Lamb HR. Will we save the homeless mentally ill? *Am J Psychiatry*. 1990;147:649–651.
13. Goldman HH, Manderscheid RW. Chronic mental disorder in the United States. In: Manderscheid RW, Barrett SA, eds. *Mental Health, United States, 1987*. Rockville, Md: US Dept of Health and Human Services, National Institute of Mental Health; 1987:1–11.
14. Arce AA, Tadlock M, Vergare MJ, Shapiro SH. A psychiatric profile of street people admitted to an emergency shelter. *Hosp Community Psychiatry*. 1983;34:812.
15. Bassuk EL, Rubin L, Lauriat A. Is homelessness a mental health problem? *Am J Psychiatry*. 1984;141:1546–1549.
16. Burt MR, Cohen BE. *America's Homeless: Numbers, Characteristics, and Programs That Serve Them*. Washington, DC: Urban Institute Press; 1989:36.
17. Spitzer RL, Williams JBW, Gibbon M, First MB. *Structural Clinical Interview for DSM-III-R*. Washington, DC: American Psychiatric Press; 1990.
18. Rossi PH, Wright JD, Fisher GA, Willis G. The urban homeless: estimating composition and size. *Science*. 1987;235:1336–1341.
19. Susser E, Struening EL, Conover S. Psychiatric problems in homeless men. *Arch Gen Psychiatry*. 1989;46:845–859.
20. Goldstein MJ. Further data concerning the relation between premorbid adjustment and paranoid symptomatology. *Schizophr Bull*. 1978;4:236.
21. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13:261–275.
22. Caton CLM. *The Community Care Schedule*. New York, NY: New York State Psychiatric Institute; January 1989.
23. Pai S, Kapur RL. The burden on the family of a psychiatric patient: development of an interview schedule. *Br J Psychiatry*. 1981; 138:332–335.
24. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NY: John Wiley & Sons Inc; 1989.
25. Caton CLM, Eagle P, Felix A, Shrout P. Family factors in the genesis of homelessness. Presented at the 120th Annual Meeting of the American Public Health Association; November 11, 1992; Washington, DC.
26. Sosin MR, Grossman S. The mental health system and the etiology of homelessness: a comparison study. *J Community Psychology*. 1991;19:337–351.
27. Belcher JR. On becoming homeless: a study of chronically mentally ill persons. *J Community Psychology*. 1989;17:173–185.
28. Osher FC, Kofoed LL. Treatment of patients with psychiatric and psychoactive substance abuse disorders. *Hosp Community Psychiatry*. 1989;40:1025–1030.
29. Caton CLM, Wyatt RJ, Felix A, Grunberg J. Follow-up of chronically homeless mentally ill men. *Am J Psychiatry*. 1993;150: 1639–1642.
30. Abram KM, Teplin LA. Co-occurring disorders among mentally ill jail detainees. *Am Psychologist*. 1991;46:1036–1045.
31. Steadman HJ, McCarty DW, Morrissey JP. *The Mentally Ill in Jail: Planning For Essential Services*. New York, NY: Guilford Press; 1989.

AIDS, HIV Infection, and Illicit Drug Use within Inner-City Families and Social Networks

ABSTRACT

Objectives. Drug use is commonly depicted and treated as an individual problem. This study describes the extent of drug use, human immunodeficiency virus (HIV) infection, and acquired immunodeficiency syndrome (AIDS) among drug users' sexual partners and family and household members in order to broaden considerations of risk.

Methods. Social network charts and structured interviews were administered to 126 women (predominantly African American and Latino) enrolled in methadone treatment. The charts elicited the prevalence of drug use and HIV infection among subjects' family and household members.

Results. Drug use and HIV/AIDS permeated subjects' sexual, familial, and household relationships. More than half of the women who resided with a sexual partner reported that their partners currently used drugs. Almost one third of the subjects' siblings were drug users. Of the 715 total siblings (all subjects plus their siblings), 69 (9.7%) were known to be HIV positive or to have an AIDS diagnosis.

Conclusions. The sexual, familial, and household expressions of drug use underscore the notion that drug use and attendant risks, including but not limited to HIV infection, might be usefully viewed and treated as an inter- and intracommunity problem rather than as an exclusively individual one. (*Am J Public Health.* 1994;84:271–274)

Anitra Pivnick, PhD, Audrey Jacobson, MD, MPH, MA, Kathleen Eric, RN, MPH, Lynda Doll, PhD, and Ernest Drucker, PhD

Introduction

The identification of illicit drug use, first intravenous heroin and cocaine and more recently crack (a smokable cocaine derivative), as behavior associated with increased risk of human immunodeficiency virus (HIV) infection has resulted in a body of research that includes the study of individual behaviors in the transmission of HIV in addition to factors that are exclusively biological.[1–7] Because individual human behavior originates in and is perpetuated through social activities and cultural contexts, studies of high-risk behaviors such as needle sharing and unprotected sex suggest the need to further broaden our notions of risk to include the social and cultural aspects of drug use in people's lives.[8,9] This understanding might then inform the development of culturally appropriate and socially effective interventions that address the problems of drug use, HIV infection, the acquired immunodeficiency syndrome (AIDS).

This paper represents an effort to characterize several important aspects of the social environment in which many drug-using women live (in this case, poor methadone patients in the Bronx). Epidemiological studies and case reports of the prevalence of HIV/AIDS in this area,[10] as well as the prevalence of HIV infection in the methadone program from which our subjects were drawn,[5] suggested the value of a closer examination of these women's family and social networks. This study was undertaken to explore the relationships between drug use, HIV infection, and social networks, including family members, sexual partners, household members, persons at risk of HIV infection, and friends.

The study was conducted from 1988 to 1990 in a methadone maintenance clinic administered by a teaching hospital in the Bronx, New York. The methadone program, which serves 700 patients, includes the administration of methadone (a substitution treatment for people with histories of opiate addiction), counseling services, an HIV testing and counseling program, and primary health care.

The fieldwork for the study was conducted through the Women's Center, a peer support program operating in the drug treatment program. Created in 1987, the Women's Center program includes support groups that provide women in methadone treatment with an opportunity to speak freely in a group setting about drug use, HIV infection, and personal matters. The program also serves as a field station for research related to social and cultural aspects of drug use, HIV infection, and HIV transmission.

Methods

The study is based on in-depth interviews with 126 women in methadone treatment. The first 30 subjects were participants in the weekly Women's Center peer support group. The remaining 96 were referred from the methadone clinic population by the original 30 subjects, by

Anitra Pivnick, Audrey Jacobson, Kathleen Eric, and Ernest Drucker are with the Department of Epidemiology and Social Medicine, Montefiore Medical Center, Bronx, NY. Lynda Doll is with the Division of HIV/AIDS, National Center for Infectious Diseases, Atlanta, Ga.

Requests for reprints should be sent to Anitra Pivnick, PhD, 3 W 87th St, New York, NY 10024.

This paper was accepted July 19, 1993.

TABLE 1—Sample Characteristics

	Sample (n = 126)
Median age, y	35
Median time in methadone treatment, y	2
Race/ethnicity, %	
Latino	62
African American	30
Caucasian	8
Grade level of education, %	
<9	12
9–11	53
≥12	35
Marital status, %	
Married/separated	59
Single/divorced/widowed	41
Source of income, %	
Public assistance	83
Employment	2
Other	15

TABLE 2—Subjects' Household Membership (n = 126)

	% of Sample
Subject, male partner, and children younger than 18 y[a]	31.0
Subject and partner[a]	19.0
Subject and children younger than 18 y[a]	23.0
Subject alone	11.1
Subject and relatives	9.5
Subject and nonrelatives	6.3

[a]Household may also include coresiding mothers, in-laws, siblings, nieces, nephews, and drug associates.

TABLE 3—Current Drug Use of Subjects and Coresiding Sexual Partners (n = 63 Couples)

Drug/Use by Partners	Use of Drug by Subject: No	Use of Drug by Subject: Yes	P[a]
Heroin			NS
No	32	11	
Yes	16	4	
Cocaine			.01
No	37	10	
Yes	7	9	
Crack			.001
No	46	8	
Yes	0	9	

[a]Significance of association between subject's and partner's use of drug.

clinical staff, and by methadone counselors. After their interviews, women not previously in the support group were invited to join as part of their HIV prevention education.

Interviews lasted for 2 hours. Each began with the administration of a kin-chart sociogram (charting kinship and social relations), followed by questions covering demographic items, history of drug use and reproduction, and HIV status. In addition, in completing a mother–child coresidence history, the interviewer inquired about the periods of time during which each woman's children had and had not lived with her, the reasons for separations, a description of both the child's and the mother's relationship to the people with whom the child lived during separation, and whether custody was informal, temporary, or legal.

The charting of kinship and social relations used in this study has its origin in classical anthropological studies of kinship.[11] In this charting method, structured with the study participant as the point of reference and stratified by generation, a model is constructed of a subject's kin relations, starting with those in closest biological or social proximity and moving outward to include persons more and more distantly related.

Similarly, in this study, the investigator initiated the construction of a kin-chart-sociogram by asking each subject a sequence of questions designed to identify blood relatives and those related to the subject through marriage. Beginning with the subject's siblings and progressing through all relatives with whom the subject was personally acquainted, the chart was completed with the identification of the subject's drug-using and non-drug-using associates and friends. Once all of these individuals had been identified and their genders, ages, places of residence, and HIV status (if known) noted, they were assigned attributes. These attributes included drug use and type, friend, sexual partner, household member, member of children's household (if children were residing separately from their mother), serostatus, and knowledge of the subject's serostatus.

Data analyses involved comparisons of proportions. Chi-square tests or, in the case of small numbers, Fisher's Exact Tests were used to determine the two-tailed statistical significance of associations. An alpha level of .05 was used as the indicator of statistical significance. The data were analyzed with SAS software.

Results

Population Characteristics

The population characteristics of the subjects are shown in Table 1. The concentration of ethnic minorities and women on public assistance is noteworthy.

Drug use. Despite their enrollment in a methadone program, 72% of the women in the study reported current illicit drug use. Twenty-five percent of the current drug users reported polydrug use. Crack was more frequently reported to be used on a daily basis (71% of the 41 users) than were either heroin (30% of the 30 users) or injected cocaine (28% of the 43 users), $\chi^2 = 18.808$, 2 *df* ($P < .001$).

HIV serostatus. Eighty-four percent of the 126 women reported having been tested for HIV infection. Of these, 48% stated that they were HIV positive, 43% stated that they were HIV negative, and 9% stated that they did not know their test results.

Household composition. Subjects reported six principal household membership patterns (see Table 2). Most of the women lived with a family member or members (including children and sexual partners), and half lived with at least one of their children. Only 11% lived alone.

Sexual partners. Half of the women reported living with a sexual partner. Most of these (82.5%) lived with a male partner, and 17.5% lived with a female partner.

Mothers and children. At the time of the interview, 106 women had 250 children under 18 years of age, half of whom were living with their mothers.

The Social Context of Drug Use

Subjects and coresiding sexual partners. Seventy percent of the 63 women in residence with a sexual partner reported that they regularly used illicit drugs and that 57% of their partners did so. Although only a minority of the partners (<25%) used crack or other forms of cocaine, significant positive associations were found between women's use of crack and other forms of cocaine and that of their partners. This pattern differed from that of heroin users; no significant association was found between women's current use of heroin and that of their coresiding sexual partners (see Table 3).

Children and household drug use. Drug use was common in the households with children under the age of 18: of the 142 children who resided with their mothers, 64.8% lived with a mother who reported regular use of illicit drugs, 82.4%

lived with at least one parent who used drugs, and 86.6% lived in households where at least one adult household member was a drug user.

History of drug use in subjects' families of origin. Histories of drug and alcohol abuse were reported in a minority of the women's parents. Among 240 known parents, 7.17% were reported to have abused alcohol and 7.9% were reported to have histories of illicit drug use.

Illicit drug use was widespread among subjects' siblings. One hundred twenty-six women in the study reported a total of 589 siblings (mean = 4.7; range = 0 to 20). Of the 119 women who had one or more siblings, 70.6% reported that at least one had a history of illicit drug use.

Of the 589 siblings, 30.6% were reported to have a history of drug abuse, not including alcohol. Eighty-five percent of these drug-using siblings were reported to have used heroin, some form of cocaine, or both. Only 6.3% of the subjects reported having a sibling who abused alcohol.

HIV Infection

Shared knowledge of positive serostatus. Among the 23 HIV-positive women with coresiding sexual partners (17 male partners and 6 female partners), 87% reported having informed their sexual partners of their serostatus. Seventy-five percent of the 40 HIV-positive women whose mothers were alive at the time of the interview reported having informed their mothers of their serostatus. Of the 27 HIV-positive women who knew their fathers and whose fathers were alive at the time of the interview, 48.1% reported having informed their fathers of their serostatus. Of the 47 HIV-positive women with siblings, 63.8% had shared their serostatus with at least one of their siblings and 34.0% had shared their serostatus with all of their siblings.

Serostatus and drug use. HIV-positive (n = 50) and HIV-negative (n = 46) women were compared in terms of type of drug used and frequency of use. Significantly more HIV-positive women (46%) than HIV-negative women (24%) reported current use of crack (chi-square test; $P < .03$). No other significant differences in drug use patterns were noted.

HIV infection in households. The presence of HIV infection in this population can be described in the context of the subjects' households. Forty-five percent of the 119 households were reported to have at least one member (including our subjects) who was HIV infected. Among these 54 households, 33.2% of the 202 household members were reported to be HIV infected (43 households included 1 HIV-infected person, 9 households included 2, and 2 households included 3).

Seventy-one children younger than 18 years lived in 50% of the 54 households with HIV-positive members; 7% of these 71 children were reported to be HIV infected. All were AIDS diagnosed.

HIV infection among siblings. Of the 715 total siblings (all subjects plus their siblings), 9.7% were reported to be HIV infected.

HIV infection, AIDS, and death by violent means. The subjects reported lifetime knowledge of 105 people living with HIV infection, 63 AIDS deaths, and 44 deaths by violent means among a total of 5053 known relatives of all ages. (Subjects were asked to identify family members who were HIV infected and those who were diagnosed with or had died of AIDS. In addition, in the course of constructing the kinchart-sociograms, women volunteered information about relatives who had died by violent means, including accidental death by fire, drowning, electrocution, or motor vehicle accident; homicide; drug-related deaths [overdose]; and suicide.) These HIV infections, AIDS deaths, and violent deaths were reported from among subjects, siblings, husbands, lovers, children, nieces, nephews, aunts, uncles, siblings-in-law, cousins, ex-husbands, ex-siblings-in-law, and stepparents.

Discussion

Drug Use

All data regarding drug use of the respondent or of her family or partner relied entirely on her own report, a clear limitation of the study. We consider it likely that such use was underreported rather than exaggerated. Comparable rates, at least of cocaine use, have been previously described among the same clinic population.[12]

Within subjects' families of origin, almost one third of the subjects' siblings were drug users. The relatively low number of drug users among parents (<8%) in comparison with the high number of their offspring who were drug users (>42%) suggests a marked increase in drug use between generations.

The social character of drug use is further evident in the number of households in which one or more member was a drug user and the number of children growing up in these households. The widespread presence of drug-using adults in households with minor children supports a perception of drug use and related behavior as normative.[13,14]

Drug use among sexual partners illustrates its relational character. All of the women who lived with a crack-using partner used crack themselves. The positive association of women's cocaine and crack use with their partners' use of the same drugs suggests some of the difficulties drug-using women may have when they try to stop using cocaine or crack.

HIV Infection

The familial and residential impact of HIV infection among African-American and Latino inner-city residents is reflected in our data in the number of subjects, subjects' relatives, and household members infected with HIV. The generational character of the HIV epidemic in New York City is reflected in the high number of infected siblings (largely of the subjects' generation).

The number of HIV-positive women who had shared knowledge of their serostatus with sexual partners and other family members suggests the persistence of familial intimacy. It also emphasizes the need for supportive services, not only for those who are HIV infected but for their family members as well.

The impact of the AIDS epidemic is reflected in our data through the number of persons reported to be HIV infected, which, according to the reports of our subjects, far exceeds the cumulative number of deaths attributed to violence throughout their lifetime.

Conclusions

The development of risk reduction strategies might profitably involve individuals or groups from drug users' social environments and communities, including active drug users and their organizations (as in Europe and Australia), for the purpose of community education, prevention, and support.[15,16] Meaningful employment may also contribute to decreased drug use by providing alternate social networks and enhanced self-esteem.

Finally, these findings encourage the provision of intensive and extensive primary prevention services, improved access to health care, and drug treatment options that incorporate the social nature of drug use and its expression in sexual relationships, households, families, and communities. □

Acknowledgments

The research for this article was supported by the Centers for Disease Control (H64/CCH20071408).

We would like to thank the women who participated in the study, the Women's Center peer facilitators, and the staff of the Montefiore Methadone Maintenance Treatment Program.

References

1. Marmor M, Des Jarlais D, Cohen H, et al. Risk factors for infection with human immunodeficiency virus among intravenous drug abusers in New York City. *AIDS.* 1987;1:39–44.
2. Centers for Disease Control. Continuing increase in infectious syphilis—United States. *MMWR.* 1988;37:35–38.
3. Chaisson RE, Bacchetti P, Osmond D, Brodie E, Sande MA, Moss AR. Cocaine use and HIV infection in intravenous drug users in San Francisco. *JAMA.* 1989;261:561–584.
4. Friedland G. AIDS and IV drug abuse. In: Kaslow et al., eds. *The Epidemiology of AIDS.* New York, NY: Oxford University Press; 1989.
5. Schoenbaum EE, Hartel D, Selwyn PA, et al. Risk factors for human immunodeficiency virus infection in intravenous drug users. *N Engl J Med.* 1989;321:874–879.
6. Schoenbaum EE, Hartel D, Friedland G. HIV infection and intravenous drug use. *Curr Opinion Infect Dis.* 1990;3:80–93.
7. Drucker E. AIDS and addiction in New York City. *Am J Drug Alcohol Abuse.* 1986;12:165–181.
8. Grund JP, Kaplan CD, Adrianns NFP. Needle sharing in The Netherlands: an ethnographic analysis. *Am J Public Health.* 1991;81:1602–1607.
9. Pivnick A, Jacobson A, Eric K, Mulvihill M, Hsu MA, Drucker E. Reproductive decisions among HIV-infected drug-using women: the importance of mother-child coresidence. *Med Anthropol Q.* 1991;2:153–169.
10. Drucker E, Vermund S. Estimating population prevalence of human immunodeficiency virus infection in urban areas with high rates of intravenous drug use: a model of the Bronx in 1988. *Am J Epidemiol.* 1989;130:133–142.
11. Evans-Pritchard EE. *The Nuer.* Oxford, England: Clarendon Press; 1940.
12. Hartel D, Schoenbaum EE, Selwyn PA, et al. Temporal patterns of cocaine use and AIDS in intravenous drug users in methadone maintenance (MM). Presented at the Fifth International Conference on AIDS, June 1989; Montreal, Quebec, Canada.
13. Bauman PS, Dougherty F. Drug addicted mothers and their children's development. *Int J Addict.* 1983;18:291–302.
14. Deren S. Children of substance abusers: a review of the literature. *J Subst Abuse Treatment.* 1986;3:77–94.
15. Grund JP, Stern LS, Kaplan CD, et al. Drug use contexts and HIV consequences. *Br J Addict.* 1992;87:381–392.
16. Wijngaart GF. *Competing Perspectives on Drug Use: The Dutch Experience.* Amsterdam, The Netherlands: Swets and Zeitlinger; 1991:40.

Injecting Risk Behavior among Drug Users in Amsterdam, 1986 to 1992, and Its Relationship to AIDS Prevention Programs

Erik J. C. van Ameijden, MS, Anneke (J) A. R. van den Hoek, MD, PhD, and Roel A. Coutinho, MD, PhD

ABSTRACT

Objectives. Serial, cross-sectional trends in injecting risk behavior were studied among drug users from 1986 to 1992.

Methods. From a cohort study in Amsterdam, 616 intake visits of drug users who had injected in the 6 months preceding intake were selected.

Results. The proportion of drug users who reported borrowing and lending used injection equipment and reusing needles/syringes (in the previous 6 months), continuously declined from 51% to 20%, from 46% to 10% and from 63% to 39%, respectively. In multivariate analysis, it appeared unlikely that a selective recruitment of participants over time was responsible for these trends. Participants, recruited later in time, had been previously tested for human immunodeficiency virus (HIV) more often, had received daily methadone less often, and had obtained a higher proportion of new needles via exchange programs. Indications were found that (1) voluntary HIV testing and counseling leads to less borrowing, lending, and reusing equipment; and (2) obtaining needles via exchange programs leads to less reusing needles/syringes. It appeared that nonattenders of methadone and exchange programs have reduced borrowing and lending to the same extent as attenders.

Conclusions. Methodologically, evaluating specific measures is difficult. However, the combination of various preventive measures in Amsterdam is likely to be responsible for the observed decrease in injecting risk behavior. (*Am J Public Health.* 1994;84: 275–281)

Introduction

In several studies among injecting drug users in Amsterdam, prevalence of the human immunodeficiency virus (HIV) was estimated to be 30% and appeared to be stable from 1986 to 1992.[1–4] In a cohort study in Amsterdam, HIV incidence initially declined from 10 per 100 person-years in 1986 to 4 in 1987 and remained stable up to 1991.[5]

However, trends in HIV prevalence and incidence are difficult to interpret. First, HIV prevalence is the result of many complex mechanisms—for example, the geographical migration of injecting drug users; the high rates of initiation and cessation of injecting drug use; the high acquired immunodeficiency syndrome (AIDS) and non-AIDS mortality, both of which are largest among HIV positives[6]; and the HIV incidence. Second, HIV incidence in itself is also influenced by various factors—for example, HIV prevalence, mixing patterns,[7] and risk behavior; thus, to estimate HIV incidence, large cohorts of drug users are necessary, and because of specific methods of recruitment and relatively high dropout rates, generalization of incidence is difficult.

Therefore, to monitor the HIV epidemic among injecting drug users, it is important to study trends in risk behavior. And among this population, the sharing of injecting equipment is the predominant route by which HIV is transmitted. Moreover, if the effect of a prevention program aimed at reducing injecting risk behavior is to be evaluated, the extent to which the sexual transmission of HIV influences the prevalence and incidence of the virus among injecting drug users must also be considered.

In Amsterdam, easily accessible methadone programs and a large-scale needle/syringe exchange program have been implemented within the concept called harm reduction[8] (see Methods). In the cohort study in Amsterdam, early results (follow-up was conducted from 1986 to 1988) indicated that the borrowing of used injection equipment decreased only as a function of the number of follow-up visits but not at intake visits over time.[9] Therefore, this risk reduction was attributed mainly to participation in the study. Several other studies have shown that injecting drug users are capable of reducing risk behavior,[10,11] but only a few have presented data on long-term trends.[12,13]

The present study is an update of the Amsterdam cohort study mentioned above. In that same study, only data measured at intake visits were used to determine serial, cross-sectional trends in injecting risk behavior from 1986 to 1992 among participants who, at entry, reported current injecting. The high-risk behaviors studied were borrowing and lending used injection equipment, and reusing needles/syringes. We also extended the present study to evaluate the potential protective effects on risk behavior of HIV antibody-test counseling, methadone treatments, and the obtaining of new needles via exchange programs.

The authors are with the Municipal Health Service, Department of Public Health and Environment, Amsterdam, the Netherlands.

Requests for reprints should be sent to Erik van Ameijden, MS, Municipal Health Service, Department of Public Health and Environment, PO Box 20244, 1000 HE Amsterdam, the Netherlands.

This paper was accepted September 20, 1993.

Methods

AIDS Prevention among Drug Users in Amsterdam

The specific purpose of the Amsterdam harm reduction approach is to create a situation that greatly reduces the risk of addicts harming themselves or their environment.[8] Within this context, three main measures were established to slow down the spread of HIV: (1) a needle and syringe exchange program (in 1984), (2) an information campaign targeted at drug users (in 1987), and (3) a program to distribute condoms (free of charge) among addicted prostitutes (in 1987).

The goal of the exchange program, begun through an initiative of the "junkiebond" (drug users organization) to prevent the spread of hepatitis B, is to prevent HIV infections by promoting the single use of needles and syringes and ensuring the availability of new needles. Equipment can be exchanged at methadone posts and other drug-help organizations (14 locations in 1991), and leaflets concerning safe drug use are available. To ensure a low threshold of the exchange program, no registration is necessary and high numbers of used needles and syringes can be exchanged.[14] This low threshold has most likely led to the sharp increase in distributed needles/syringes: from 100 000 per year in 1985 to approximately 1 000 000 per year in 1990, 1991, and 1992 (S. van Lieshout, personal communication).

The low-threshold methadone programs, operated by the Drug Department of the Municipal Health Service, are focused on stabilizing drug use and contacting drug users so they can receive appropriate social and medical care.[15] The illicit injection and other use of opiates or other drugs is tolerated. Although they are not primarily aimed at preventing HIV infection, these programs may contribute to AIDS prevention by lowering the frequency of injecting (methadone being a substitute for heroin) and increasing the proportion of drug users receiving educational information on AIDS.

Study Design

In December 1985, an open and ongoing cohort study among drug users was started in Amsterdam.[1,2,5,6,9] The participants for this study are recruited from methadone posts, from a sexually transmitted disease clinic for addicted prostitutes, and, to a lesser extent, by word of mouth. Participation in the study is voluntary, and informed consent is obtained. In the first 3 years, participants were seen at eight different locations. Since 1989, they have been seen at one central location, except for the addicted prostitutes, who are still interviewed at the sexually transmitted disease clinic. At intake, participants are given standardized questionnaires, which are administered by specially trained nurses; questions regarding current behavior refer to the previous 6 months.

Between December 1, 1985, and November 30, 1992, a total of 973 drug users entered this cohort study. For the present study, 616 of these users who had reported injecting during the 6 months preceding intake were selected. The only data used to determine trends in risk behavior were collected at intake visits. A large individual risk reduction has already been reported.[9] However, the validity of self-reported behavioral changes during follow-up is questionable when one considers the high and stable incidence of injecting related viral infections during follow-up.[16] In addition, injecting drug users who visited more than once appear to differ from single visitors, which might have led to a selection bias.

Trends in high-risk behaviors were determined over five periods of intake visits (1986 [n = 182], 1987 [n = 137], 1988 [n = 106], 1989/90 [n = 125], and 1991/92 [n = 66]). The smaller number of injecting drug users in the last periods is owing to a temporary cessation of recruitment of new participants between September 1990 and August 1991. The variables were divided into four sets. The first set consists of the three outcome variables that reflect current behavior (i.e., behavior measured over the 6 months preceding intake): borrowing needles or syringes already used by somebody else (yes, no), lending used needles/syringes to others (yes, no), and reusing one's own needles/syringes (mainly, not mainly). These variables do not include "frontloading" (i.e., injecting some drug solution into the front of another syringe of which the needle has been taken off in order to honestly divide drugs).[17] This behavior seems to be not very prevalent in Amsterdam.

Trends in risk behavior over time might be fully explained by a selection bias of participants over time; injecting drug users with certain characteristics that might be related to injecting risk behavior could have selectively entered the study over time. Therefore, trends in risk behavior were multivariately adjusted for differences in general characteristics (set 2) and current behavior variables (set 3), shown in Table 1. Sets 2 and 3 are distinguished because adjustment for current behavior might lead more easily to overmatching than adjustment for general characteristics; variables like age and gender cannot be influenced by interventions whereas variables for current behavior can. For example, if it is assumed that both borrowing and cocaine injecting decline in time (owing to intervention) and that cocaine injectors borrow more often than heroin injectors, adjusting the effect of year of intake visit on borrowing for frequency of cocaine injecting would lead to overmatching.

In 1989, the coding categories of the general characteristics of time living in Amsterdam, duration of injecting, and frequency of borrowing since 1980 were changed from broad categories to exact number of years or times. Only for frequency of borrowing did this result in more missing values (before 1989, 1% were missing; from 1989 on, 16% were missing.) Subjects with missing values on this variable were included in the multivariate analyses within a separate category, "missing." On all other variables, less than 1% of subjects had missing values. The current behavior variables were not subject to important changes.

The intervention variables, which may have a protective effect on high-risk behavior, form the fourth set. These variables include being previously tested for HIV (never, tested HIV positive, tested HIV negative), currently receiving methadone daily via methadone programs (yes, no), and currently obtaining new needles/syringes via exchange programs (0%, 1% to 99%, 100%).

In general, all testing was done two-tailed, and a P value of .05 was the criterion of significance. Statistical significance of trends was univariately analyzed using chi-square trend tests for categorical variables and Spearman's rank correlation tests for continuous variables. Multivariately, logistic regression was used to adjust for differences between participants over time, to determine significant and independent predictors for the outcome variables, and to investigate interactions between important variables.[18] Significance was based on the likelihood ratio statistic. The fit of the models was evaluated by calculating variable specific delta betas; no observations showed a disproportionally large influence on the odds ratios. The fit was also assessed using the Pearson goodness-of-fit test; this test was never significant. For all analyses, SPSS 4.0 was used.[19]

Results

Of the 616 injecting drug users who had injected during the 6 months preceding intake and who entered our study between 1986 and 1992, 57% were men and 43% women, with mean ages of 31.2 (SD = 5.8) and 27.5 (SD = 5.7) years, respectively. Of these current injectors at intake, 57% had their last injection 1 day ago or less, 18% had injected 2 to 7 days ago, 13% had injected 1 to 4 weeks ago, and 13% had injected 1 to 6 months ago. Sixty percent had currently injected mainly speedballs (a mixture of heroin and cocaine), and 24% had injected mainly heroin.

Unadjusted and Adjusted Trends in High-Risk Behavior

The three outcome variables showed a strong and highly significant decline per intake period (Table 1). From 1986 through 1991/92, borrowing needles/syringes decreased from 51% to 20%; lending, from 46% to 10%; and reusing, from 63% to 39% (all $P < .0001$). The concept of role separation is of epidemiological importance here since the incidence pattern of HIV depends on the proportion of injecting drug users who only borrow or lend and on the porportion of those who do both.[20] In 1986, 36% reported neither borrowing nor lending, 14% reported only lending, 18% reported only borrowing, and 32% reported both risk behaviors. In 1991/92, these figures were 77%, 3%, 13%, and 7%, respectively.

Table 1 also shows trends in general characteristics and current behavior variables per intake period. Per intake period, both the mean age and the proportion of men significantly increased, whereas the frequency of borrowing since 1980; the proportion of injecting drug users currently prostituting; and the proportions of users using amphetamines, barbiturates, and tranquilizers orally all show decreases.

Unadjusted and adjusted trends in the outcome variables were calculated by entering the period of intake visit continuously into logistic regression models in three ways: (a) unadjusted (bivariately), (b) adjusted for general characteristics (set 2), and (c) additionally adjusted for current behavior variables (sets 2 and 3). It appears that the odds ratios reflecting the decrease in borrowing, lending, and reusing needles per period of intake visit hardly changed after adjustment (model a compared with models b and c, Table 2). Therefore, the decrease in high-risk behavior was probably not caused by selective participation over time.

TABLE 1—Trends in Outcome Variables, General Characteristics, and Current Behavior Variables, among 616 Intake Visitors Who Had Injected Drugs during the Previous 6 Months

	Period of Intake Visit					
	1986 (n = 182)	1987 (n = 137)	1988 (n = 106)	1989/90 (n = 125)	1991/92 (n = 66)	*P*
Outcome variables[b] (set 1)						
Borrowing needles/syringes	51[a]	51	45	38	20	***
Lending needles/syringes	46	30	34	14	10	***
Reusing needles/syringes	63	54	51	42	39	***
General characteristics (set 2)						
Mean age, y	28.7	29.2	28.2	31.2	32.2	***[c]
Male	45	58	64	62	67	***
Nationality						
Dutch	66	58	63	54	67	
German	23	23	19	25	9	
Living >5 years in Amsterdam	53	61	58	45	61	
Duration of injecting						
<1 year	7	7	9	6	6	
>5 years	75	75	65	71	63	
Frequency of borrowing since 1980[d]						
<10 times	55	64	55	71	77	*
100 or more times	12	13	12	15	9	
Last moment of injection						
<2 days ago	59	58	55	51	62	
1–6 months ago	12	11	13	14	14	
Prostitution[b]	43	31	34	37	24	*
Current behavior variables[b] (set 3)						
Injecting >1 time/day	53	45	40	55	49	
Type of drug mainly injected						
Heroin alone	23	22	20	28	22	
Cocaine alone	9	10	11	7	8	
Heroin and cocaine	57	61	61	58	66	
Amphetamine	8	7	5	3	2	
Heroin use, nonparenteral	34	45	51	38	48	
Cocaine use, nonparenteral	25	34	40	23	36	
Amphetamines use, orally	14	9	8	2	3	***
Barbiturates use, orally	22	10	19	12	6	**
Tranquilizers use, orally	66	61	50	36	33	***
Alcohol use, >4 glasses/day	20	23	18	21	12	
Having a currently injecting steady sex partner	25	22	19	24	17	

[a]In percentages, except for age.
[b]Reported over the 6 months preceding intake.
[c]By Spearman's rank test.
[d]36 had missing values (see Methods).
*$P < .05$, **$P < .01$, ***$P < .001$ (by chi-square trend test).

Evaluation of Specific Interventions

The variables concerned with potentially protective interventions also showed trends. Participants recruited later in time had been previously tested for HIV more often, received daily methadone less often, and obtained a higher proportion of new needles via exchange programs (all $P < .0001$, Table 3). The decrease in those receiving methadone is owing to a selective recruitment of participants at the methadone posts early in time. It may be that the decrease in risk behavior is related to these trends in intervention variables; indeed, in bivariate logistic regression analyses (unadjusted), injecting drug users who obtained all needles via exchange programs and those who had previously been tested for HIV (especially those whose test results were HIV positive) reported less high-risk behavior: they borrowed, lended, and reused needles significantly less often (Table 4). Receiving daily methadone was not strongly associated with either outcome.

For each outcome, a multivariate logistic regression model was constructed to determine independent predictors among the intervention variables and year of in-

TABLE 2—Logistic Regression Analysis: Unadjusted and Adjusted Trends in Outcome Variables, among 616 Intake Visitors Who Had Injected Drugs during the Previous 6 Months

	Effect of Intake Period[a] on		
	Borrowing	Lending	Reusing Needles
Unadjusted OR[b]	0.77**	0.61**	0.78**
Adjusted OR for set 2, general characteristics	0.79**	0.58**	0.82*
Adjusted OR for sets 2 and 3, general characteristics and current behavior variables	0.80*	0.57**	0.81*

[a]Continuous into model (the odds ratios correspond with every increase of 1 unit in intake period; e.g., using 1987 as a reference, the odds of borrowing [unadjusted] in 1988 are 0.77 and in 1989/90 are 0.77^2.
[b]OR = odds ratio.
*$P < .01$, **$P < .001$ (likelihood ratio test).

TABLE 3—Trends in Intervention Variables among 616 Intake Visitors Who Had Injected Drugs during the Previous 6 Months

	Period of Intake Visit					
Intervention Variables (set 4)	1986 (n = 182), %	1987 (n = 137), %	1988 (n = 106), %	1989/90 (n = 125), %	1991/92 (n = 66), %	P
Being previously tested for HIV						
Never tested	100	100	95	74	64	*
HIV positive	0	0	2	10	9	
HIV negative	0	0	3	16	27	
Receiving daily methadone[a]	85	86	70	67	62	*
Obtaining new injecting equipment via needle exchange programs[a]						
0%	51	41	25	15	20	*
1–99%	18	15	26	24	27	
100%	31	44	50	60	53	

Note. $P < .001$ (chi-square trend test).
[a]Reported over the 6 months preceding intake.

take visit, adjusted for general characteristics. (Additionally adjusting for current behavior variables yielded very similar results.) With regard to the outcomes of borrowing and lending, comparable results were found (Table 4). The likelihood of borrowing and lending decreased with a later period of intake visit and with having been previously tested for HIV, both variables bearing a stronger relation to lending than to borrowing. Multivariately adjusted, the relations between pbx-exchanging and the outcomes of borrowing and lending were weaker than they were unadjusted and were nc longer significant. Further analyses revealed that these relations weakened because of adjustment for the variable period of intake visit only.

Different results were obtained with regard to the outcome of reusing needles/syringes. Independent and significant predictors were having received an HIV-positive test result and exchanging needles/syringes frequently. However, the effect of period of intake visit was much weaker and no longer significant; this is probably explained by the increasing proportion of injecting drug users who frequently exchange over time.

Both the outcome variables and the intervention variables show statistically significant trends per intake period. It appears that these variables are also significantly interrelated except for the associations between receiving daily methadone and the three outcome variables or exchanging (Table 5). Therefore, severe problems of multicollinearity arise. For instance, adjusting the relationship between exchanging and borrowing or lending for period of intake visit might produce overmatching; however, one should adjust for calendar time to control for the potential presence of other (secular) trends (e.g., other preventive measures taken or a general spread of knowledge of safe injection methods).

To gain more insight into the mechanisms of the observed risk reduction, trends were compared in risk behavior among injecting drug users who did and did not receive daily methadone and who obtained syringes 100% and less than 100% via exchange. It appears that the proportion of users who reported current borrowing and lending decreased over time in all subgroups. Whether the magnitude of trends differed between the subgroups was then formally tested by evaluating if first-order interaction terms (i.e., between exchanging or receiving methadone and period of intake visit) could significantly improve a logistic regression model also containing the main effects. None of these interaction terms could do so (likelihood ratio test: $P > .11$).

Injecting drug users who obtained 100% of new needles/syringes via exchanging reported as early as 1986 that they were reusing their needles less often than those users who were exchanging them less. The relationship between reusing and exchanging became even stronger in time. This difference in the magnitude of trends (i.e., in subgroups based on strata of exchanging) was statistically significant (likelihood ratio test: $P = .04$). In 1986, injecting drug users who received daily methadone treatments also reported reusing needles less often than those who received methadone irregularly or not at all. However, the relation between reusing and receiving methadone disappeared in 1991/92. This difference in magnitude of trends was of marginal statistical significance (likelihood ratio test: $P = .08$).

Discussion

Among intake visitors in our cohort study of injecting drug users in Amsterdam, a substantial and continuous risk reduction occurred between 1986 and 1992. The proportion reporting current borrowing declined from 51% to 20%; lending from 46% to 10%; and reusing, from 63% to 39% (all $P < .0001$). However, there were some potential sources of bias in these trends.

First, participants showing a high level of risk behavior may have been selectively recruited earlier in time in our study. Therefore, we adjusted these trends multivariately for differences in eight general characteristics; the same decline in risk behavior was found. A possible drawback to this method may have been that injecting drug users differed in characteristics for which no information was available. However, even after addi-

tionally adjusting for nine current behavior variables, which theoretically could lead to overmatching, we found the same magnitude of decline in borrowing, lending, and reusing needles.

Second, we selected only drug users who had injected in the preceding 6 months. Assuming that, under the influence of the HIV epidemic, drug users initiate injection less often or are more likely to cease injecting, the decline in risk behavior among all drug users in Amsterdam would be underestimated.

Third, the prevalence of borrowing and lending may be underestimated owing to memory loss and the deliberate giving of socially desirable answers. However, underreporting would have had to increase systematically in time to explain the decrease in risk behavior. When we adjusted for having been tested for HIV and for participation in methadone and needle exchange programs—assuming that these interventions are markers for social desirability because they reflect more knowledge of risky behavior—the decline in borrowing and lending was still statistically significant. However, one can never rule out this type of bias since the norms against sharing in the entire injecting drug culture may have been strengthened.

Finally, the outcomes of borrowing and lending are only roughly measured. For example, it is likely that injecting drug users who borrowed injecting equipment that had already been used borrowed less often in the previous 6 months, borrowed less often from those testing HIV positive (e.g., they borrowed only from their HIV-negative steady partner), or adequately cleaned the needles and syringes more often. Given such changes, the decline in risk behavior would be underestimated.

The overall direction of these potential sources of bias is unknown. However, the observed risk reduction is in accordance with the stable HIV prevalence in Amsterdam and the decline in HIV incidence in our cohort.[5] We want to stress that, in 1992, 20% of injecting drug users still reported current borrowing and that the HIV incidence within our cohort was estimated to be 3 per 100 person-years. This indicates the need to improve the AIDS prevention campaign among injecting drug users in Amsterdam.

In the present study, we tried to evaluate the effects of three interventions on injecting risk behavior. Because of two general methodological problems, however, our results should be interpreted with caution. First, secular trends may be present. For instance, before AIDS prevention measures had been established in Amsterdam, it was already public knowledge that HIV could be transmitted parenterally. In addition, shortly after the start of specific programs for drug users, a national large-scale media campaign was executed in the Netherlands. Therefore, causality is difficult to prove. However, by multivariately adjusting the relationships between high-risk behavior and the intervention variables for year of intake period, we have (in part) adjusted for such secular trends. Second, in the present cross-sectional study, cause and effect cannot be distinguished. In Amsterdam, both HIV testing/counseling and attending exchange programs are voluntary. Thus, injecting drug users who were strongly motivated to reduce risk behavior or had already reduced it might have been attracted to these two interventions. On the other hand, the prevention programs can provide the means by which persons who are trying to avoid risk are in fact able to do so.

TABLE 4—Logistic Regression Analyses: Associations between the Borrowing, Lending, and Reusing of Needles and Syringes and the Intervention Variables, among 616 Intake Visitors Who Had Injected Drugs during the Previous 6 Months

	Outcome Variable					
	Borrowing		Lending		Reusing Needles	
	Unadjusted OR	Adjusted[a] OR	Unadjusted OR	Adjusted[a] OR	Unadjusted OR	Adjusted[a] OR
Intervention variables						
Being previously tested for HIV						
Never tested	1	1	1	1	1	1
HIV positive	0.14**	0.18*	0.002	0.003	0.23**	0.29*
HIV negative	0.47*	0.78	0.053**	0.124*	0.67	1.17
Receiving daily methadone[b]	1.00	0.87	1.35	1.11	0.75	0.65
Exchanging[b]						
0%	1	1	1	1	1	1
1–99%	0.96	0.99	1.03	1.21	0.60*	0.60*
100%	0.69*	0.86	0.61*	0.78	0.28***	0.29***
Period of intake visit[c]	0.77***	0.83*	0.61***	0.64***	0.78***	0.89

[a]Odds ratio (OR) derived from a logistic regression model including the intervention variables, period of intake visit, and general characteristics (associations with general characteristics not shown).
[b]Reported over the 6 months preceding intake.
[c]Continuously into the model.
*$P < .05$, **$P < .01$, ***$P < .001$ (likelihood ratio test).

TABLE 5—Spearman's Rank Correlations between Period of Intake Visit, Outcome Variables, and Intervention Variables, among 616 Intake Visitors Who Had Injected Drugs during the Previous 6 Months

	Intake Period	Borrowing	Lending	Reusing Needles	Previously Tested	Received Methadone Daily
Outcome variables						
Borrowing needles/syringes	−.17***					
Lending needles/syringes	−.28***	.45***				
Reusing needles/syringes	−.17***	.19***	.18***			
Intervention variables						
Previously tested for HIV[a]	.42***	−.14***	−.20***	−.10**		
Received methadone daily	−.21***	.00	.06	−.07	−.16***	
Exchanging[b]	.27***	−.09*	−.10**	−.27***	.12**	.03

[a]Ranked 0 = HIV status unknown, 1 = HIV status known.
[b]Ranked 1 = 0%, 2 = 1–99%, 3 = 100%.
*$P < .05$, **$P < .01$, ***$P < .001$ (Spearman's rank test).

Despite these potential drawbacks, we found indications that voluntary HIV antibody testing/counseling reduces high-risk behavior. This intervention was related to all three outcomes (borrowing, lending, and reusing), independent of year of intake visit (secular trends) and other potential confounders. HIV testing is very strongly related to lending; almost none of the previously tested injecting drug users lent equipment to others. Injecting drug users who knew themselves to be HIV positive reported the lowest levels of risk behavior. However, being tested and found HIV negative does not seem to give false feelings of safety; these users were less likely to borrow than those who were never tested. In a review (1991), it was concluded that, "to date, the studies among injecting drug users do not provide compelling evidence for an effect of counseling and testing on either drug or sexual risk reduction."[11] In a clinical trial no protective effect was found.[21] Other recent studies, however, indicate beneficial effects.[22–26]

One important goal of the exchange programs is to promote the single use of needles and syringes, which is beneficial for hygienic reasons in general (e.g., to prevent bacterial and fungal infections). Beyond that, such behavior is expected to prevent HIV infection because habitual single-needle users will better organize injecting drug use, will less easily interchange injection equipment by accident, and will be more reluctant to borrow needles used by somebody else. Indeed, single-needle users reported less borrowing. The results of the present study suggest that the decrease in reusing equipment was caused by the increase in exchanging equipment. There was not only a strong but also a dose-effect relationship between exchanging and reusing needles, which was independent of year of intake visit (secular trends) and other potential confounders. Furthermore, a decline in reusing needles was only found among injecting drug users who exclusively exchanged needles. As for how attending exchange programs affected borrowing and lending, the results are difficult to interpret owing to multicollinearity (see Results).

No relation was found between receiving daily methadone and either outcome. However, because of methadone programs, drug users may have stopped drug injection. This might have led to a selection bias since participants eligible for the present study were required to have been currently injecting. Another complicating factor is that, owing to the low threshold of the programs, it is relatively easy to leave and reenter the programs. As a result, almost all injecting drug users in the present study received methadone after information on HIV became available at these programs. In any case, because in 1991 70% of the estimated 6500 drug users in Amsterdam received methadone,[27] low threshold programs are able to reach the majority of drug users, to whom information on HIV could be provided and syringe exchange and HIV testing/counseling services[28] could be offered.

Surprisingly, in the multivariate analyses, the decline in borrowing and lending was independent of intervention and other variables, and we found that injecting drug users attending needle exchanges or methadone programs reduced their risk to the same extent as nonattenders. Even among injecting drug users who neither obtained new needles via the exchange nor received methadone regularly, a significant risk reduction was found (data not shown). Not only secular trends (as described above) but also "contamination of the control group" may be responsible for these findings: nonattenders had probably received information about safe injection practices from attenders, and because of the exchange programs, the availability of needles/syringes may have been increased overall. For instance, exchangers may have given new injection equipment away to others more easily. At present, exchanging may be seen as a convenient way to obtain injection equipment, which is not strongly related to a desire to reduce risk.[5,29]

These findings have important implications for studies previously performed within our cohort. The lack of convincing evidence that attending exchange programs or receiving methadone treatments has a beneficial effect on the HIV prevalence, HIV incidence, or current borrowing[5,24,30] might be owing to the two factors mentioned above. Methodologically, evaluating specific prevention activities is very difficult. Even randomized intervention trials may be subject to secular trends and contamination of the control group. In the introduction to this paper, we mentioned limitations in the monitoring of HIV prevalence and incidence. Using the incidence of hepatitis B infection as a surrogate marker may give misleading results.[16]

However, a communitywide reduction in injecting risk behavior can be demonstrated by performing serial, cross-sectional studies using standardized questionnaires. In some serial, cross-sectional studies, a decline in injecting risk behavior was observed[31–34]; in others, no risk reduction was found.[13,23,35,36] Such differences between cities or regions may give clues to the relative effectiveness of prevention programs. However, apart from preventive measures, other differences between cities (e.g., drug use culture) complicate these comparisons.

In conclusion, a substantial risk reduction in injecting risk behavior has occurred since 1986 among injecting drug users in Amsterdam. Indications have been found that (1) voluntary HIV testing and counseling leads to less borrowing, lending, and reusing of injecting equipment; and (2) obtaining needles via exchange programs leads to less reusing of equipment. It appears that nonattenders of methadone and exchange programs have reduced borrowing and lending to the same extent as attenders. Therefore, evaluating specific measures methodologically is difficult. However, the combination of various measures in Amsterdam is probably responsible for the observed decrease in injecting risk behavior. □

Acknowledgments

This study was supported by the Netherlands Foundation for Preventive Medicine, grant no. 28–1258.

The authors wish to thank H. van Haastrecht for data management and methodological comments; the nurses B. Frölich, B. Scheeringa-Troost, G. Stienstra, J. Teeuwissen, R. Lopez-Dias, and J. Bax, and Drs G. Mientjes and H. Fennema for conducting the interviews and taking blood samples; and M. ter Pelle for handling the administration.

References

1. van den Hoek JAR, Coutinho RA, van Haastrecht HJA, van Zadelhof AW, Goudsmit J. Prevalence and risk factors of HIV infections among drug users and drug-using prostitutes in Amsterdam. *AIDS.* 1988;2:55–60.
2. van Haastrecht HJA, van den Hoek JAR, Bardoux C, Leentvaar-Kuypers A, Coutinho RA. The course of the HIV epidemic among intravenous drug users in Amsterdam, the Netherlands. *Am J Public Health.* 1991;81:59–62.
3. Fennema JSA, van Ameijden EJC, van den Hoek JAR, Coutinho RA. Incorrect prediction of HIV seroprevalence among intravenous drug users contacted at the street in Amsterdam. Presented at the Seventh International Conference on AIDS, June 16–21, 1991; Florence, Italy; WC 3041.
4. Hartgers C, van Santen GW, van Haastrecht HJA, van den Hoek JAR, van den Akker R, Coutinho RA. De HIV-prevalentie onder druggebruikers die methadon verstrekt krijgen bij de drugsafdeling van de GG&GD te Amsterdam. *T Soc Gezondheidsz.* 1992;70:275–279.

5. van Ameijden EJC, van den Hoek JAR, van Haastrecht HJA, Coutinho RA. The harm reduction approach and risk factors for human immunodeficiency virus seroconversion in injecting drug users, Amsterdam. *Am J Epidemiol.* 1992;136:236–243.
6. Mientjes GH, van Ameijden EJC, van den Hoek JAR, Coutinho RA. Increasing morbidity without rise in non-AIDS mortality among HIV-infected intravenous drug users in Amsterdam. *AIDS.* 1992;6:207–212.
7. Blower S. Behaviour change and stabilization of seroprevalence in communities of injecting drug users: correlation of causation? *J AIDS.* 1991;4:920–923.
8. Buning EC, Coutinho RA, van Brussel GHA, van Santen G, Zadelhof AW. Preventing AIDS in drug addicts in Amsterdam. *Lancet.* 1986;1:1435. Letter.
9. van den Hoek JAR, van Haastrecht HJA, Coutinho RA. Risk reduction among intravenous drug users in Amsterdam under the influence of AIDS. *Am J Public Health.* 1989;79:1355–1357.
10. Des Jarlais DC, Friedman SR. HIV and intravenous drug use. *AIDS.* 1988;2(suppl 1):s65–s69.
11. Higgins DL, Galavotti C, O'Reilly KR, et al. Evidence for the effects of HIV antibody counselling and testing on risk behaviors. *JAMA.* 1991;266:2419–2429.
12. Des Jarlais DC, Abdul-Quader A, Tross S. The next problem: maintenance of AIDS-risk reduction among intravenous drug users. *Int J Addict.* 1991;26:1279–1292.
13. Longshore D, Anglin MD, Annon K, Hsieh S. Trends in self-reported risk behavior: injection drug users in Los Angeles. *J Acquir Immune Defic Syndr.* 1993;6:82–90.
14. Drion J. Jaarverslag spuitomruilprogramma 1988. Amsterdam, the Netherlands: Municipal Health Service (GG&GD); 1989.
15. Buning EC, van Brussel GHA, van Santen GW. The "methadone by bus" project in Amsterdam. *Br J Addict.* 1990;85:1247–1250.
16. van Ameijden EJC, van den Hoek JAR, Mientjes GHC, Coutinho RA. A longitudinal study on the incidence and transmission patterns of HIV, HBV, and HCV infection among drug users in Amsterdam. *Eur J Epidemiol.* 1993;9:255–262.
17. Grund J-P, Kaplan CD, Adriaans NFP, Blanken P. Drug sharing and HIV transmission risks: the practice of frontloading in the Dutch injecting drug user population. *J Psychoactive Drugs.* 1991;23:1–10.
18. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley and Sons Inc; 1989.
19. Norusis MJ. *SPSS/PC+ Advanced Statistics 4.0.* Chicago, Ill: SPSS Inc; 1990.
20. van Druten H, van Griensven F, Hendriks J. Homosexual role separation: implications for analyzing and modeling the spread of HIV. *J Sex Res.* 1992;29:477–499.
21. Calsyn DA, Saxon AJ, Freeman G Jr, Whittaker S. Ineffectiveness of AIDS education and HIV antibody testing in reducing high-risk behaviors among injecting drug users. *Am J Public Health.* 1992;82:573–575.
22. Sasse H, Salmasso S, Conti S. Risk behaviors for HIV-1 infection in Italian drug users: report from a multicenter study. *J Acquir Immune Syndr.* 1989;2:486–496.
23. Nicolosi A, Molinari S, Musicco M, Saracco A, Ziliani N, Lazzarin A. Positive modification of injecting behavior among intravenous heroin users from Milan and northern Italy 1987–1989. *Br J Addict.* 1991;86:91–102.
24. Magura S, Siddiqi Q, Shapiro J, et al. Outcomes of an AIDS prevention program for methadone patients. *Int J Addict.* 1991;26:629–655.
25. van den Hoek JAR, van Haastrecht HJA, Coutinho RA. Little change in sexual behavior in injecting drug users in Amsterdam. *J Acquir Immune Defic Syndr.* 1992;5:518–522.
26. Vanichseni S, Choopanya K, Des Jarlais DC, et al. HIV testing and sexual behavior among intravenous drug users in Bangkok, Thailand. *J Acquir Immune Defic Syndr.* 1992;5:1119–1123.
27. van Brussel GHA, van Lieshout SJM. *Jaarverslag 1991 drugsafdeling.* Amsterdam, the Netherlands: Municipal Health Service; 1992.
28. Farley TA, Cartter ML, Wassel JT, Hadler JL. Predictors of outcome in methadone programs: effect of HIV counselling and testing. *AIDS.* 1992;6:115–121.
29. Hartgers C, van Ameijden EJC, van den Hoek JAR, Coutinho RA. Determinants of borrowing used needles and of not participating in the needle exchange programme among HIV-negative drug users. *Public Health Rep.* 1992;107:675–681.
30. Hartgers C, van den Hoek JAR, Krijnen P, Coutinho RA. HIV prevalence and risk behavior among injecting drug users who participate in "low-threshold" methadone programs in Amsterdam. *Am J Public Health.* 1992;82:547–551.
31. Guydish JR, Abramowitz A, Woods W, Black DM, Sorensen JL. Changes in needle sharing behavior among intravenous drug users: San Francisco, 1986–1988. *Am J Public Health.* 1990;80:995–997.
32. Watters JK, Downing M, Case P, Lorvick J, Cheng YT, Fergusson B. AIDS prevention for intravenous drug users in the community: street-based education and risk behavior. *Am J Community Psychol.* 1990;18:587–596.
33. Robert CF, Deglon JJ, Wintsch J, et al. Behavioural changes in intravenous drug users in Geneva: rise and fall of HIV infection, 1980–1989. *AIDS.* 1990;4:657–660.
34. Käll KI, Olin RG. HIV status and changes in risk behaviour among intravenous drug users in Stockholm 1987–1988. *AIDS.* 1990;4:153–157.
35. Chaisson RE, Osmond D, Moss AR, Feldman HW, Bernacki P. HIV, bleach and needle sharing. *Lancet.* 1987;1:1430. Letter.
36. Battjes RJ, Pickens R, Amsel Z. HIV infection and AIDS risk behaviors among intravenous drug users entering methadone maintenance treatment in selected US cities. *J Acquir Immune Defic Syndr.* 1991;4:1148–1154.

24. Armon C, Kurland LT. Chronic fatigue syndrome: issues in the diagnosis and estimation of incidence. *Rev Infect Dis.* 1991; 13(s1):68–72.
25. Straus SE. History of chronic fatigue syndrome. *Rev Infect Dis.* 1991;13(s1):19.
26. Wallace PG. Epidemiology: a critical review. *Br Med Bull.* 1991;47:942–951.
27. Koos E. *The Health of Regionville: What People Thought and Did about It.* New York, NY: Columbia University Press; 1954.
28. Valdini AF, Steinhardt S, Valicanti J. A one year follow-up of fatigue patients. *Fam Pract.* 1987;4:2,103–107.
29. Diehr P, Madden CW, Cheadle A, et al. Estimating county percentages of people without health insurance. *Inquiry.* 1991;28: 413–419.
30. Susser M, Watson W, Hopper K. *Sociology in Medicine.* 3rd ed. New York, NY: Oxford University Press; 1985.
31. Kaplan HB. Social psychology of the immune system: a conceptual framework and review of the literature. *Soc Sci Med.* 1991; 33:909–923.
32. Dohrenwend BP. Sociocultural and sociopsychological factors in the genesis of mental disorders. *J Health Soc Behav.* 1975; 16:365–392.
33. Brown GW, Harris TO. *Social Origins of Depression.* London, England: Tavistock; 1978.
34. Link BG, Dohrenwend BP, Skodol AE. Socio-economic status and schizophrenia: noisome occupational characteristics as a risk factor. *Am Sociol Rev.* 1986;51:242–258.
35. Mirowsky J, Ross CE. *Social Causes of Psychological Distress.* New York, NY: Aldien; 1989.
36. Verbrugge LM. The twain meet: empirical explanations of sex differences in health and mortality. *J Health Soc Behav.* 1989; 30:282–304.

The Provision and Use of Mental Health Services in Nursing Homes: Results from the National Medical Expenditure Survey

Michael A. Smyer, PhD, Dennis G. Shea, PhD, and Andrea Streit, MHA

ABSTRACT

Data from the Institutional Population Component of the National Medical Expenditure Survey were used to provide national estimates of annual mental health service provision and use in nursing homes. In addition, the relationship between service provision and setting characteristics such as ownership, size, Medicaid certification, and chain status was examined. Although more than three quarters of residents with a mental disorder resided at a nursing home that provided counseling services, fewer than one fifth actually received any mental health services within the year. (*Am J Public Health.* 1994;84:284–287)

Introduction

Mental health services in nursing homes are in transition. The Nursing Home Reform Act of 1987 (Public Law 100-203), part of the Omnibus Budget Reconciliation Act of 1987, mandates screening and services for mentally ill elderly persons who are seeking admission to nursing homes. Even as the legislation is being implemented, estimates are needed to provide a benchmark for progress in service provision. Recent estimates of monthly service provision[1] are based on the 1985 national nursing home survey. The purpose of this report is to provide estimates of annual service provision based on the most recent national data, before implementation of the Omnibus Budget Reconciliation Act of 1987.

This article uses data from the Institutional Population Component of the 1987 National Medical Expenditure Survey[2] to examine the relationship between the annual provision of mental health services in nursing homes and the need for and use of these services among nursing home residents. Because previous work suggests that nursing home ownership affects service provision,[3] the analysis distinguishes between for-profit, not-for-profit, government ownership, and other organizational characteristics.

Methods

Detailed information regarding the methodology for the Institutional Population Component of the 1987 National Medical Expenditure Survey is provided elsewhere.[4,5] This study examined only the resident sample obtained on January 1, 1987, which is a cross-sectional sample more likely to represent a longer staying and hence slightly more ill group.[6] The sample group of facilities analyzed here excluded those for mentally retarded persons because the focus is on the provision of mental health services in nursing homes. This sample also excluded facilities that were part of retirement centers or

The authors are with the College of Health and Human Development, The Pennsylvania State University, University Park, Pa.

Requests for reprints should be sent to Michael A. Smyer, PhD, 110 South Henderson Building, Penn State University, University Park, PA 16802.

This paper was accepted May 20, 1993.

Editor's Note. The data used in this article were made available by the Inter-University Consortium for Political and Social Research. The data for *The National Medical Expenditure Survey, 1987: Institutional Population Component* were originally collected and prepared by the National Center for Health Services Research and Health Care Technology Assessment. Neither the collectors of the original data nor the Consortium bear any responsibility for the analyses or interpretations presented here.

TABLE 1—Provision of Counseling by Nursing Homes

	Total No.	% Providing Some Counseling
All nursing homes	21 095	59.3
Ownership		
Not for profit	4 589	54.1
For profit	15 844	59.8
Government	682	82.7
Bed capacity		
3–49	9 204	40.0
50–99	5 318	67.2
100–149	4 217	80.5
150 or more	2 356	78.8
Medicaid SNF certification		
No SNF beds	13 514	50.3
1–49% SNF beds	2 216	69.6
50–99% SNF beds	1 423	82.5
100% SNF beds	4 742	74.8
Facility chain status		
Part of chain	9 721	67.1
Independent nursing home	11 374	52.7

Note. SNF = skilled nursing facility.
Source. Compiled from data from the *National Medical Expenditure Survey, 1987: Institutional Population Component.*[2]

TABLE 2—Types of Personnel Providing Mental Health Services

	Percentage of Nursing Homes Reporting Services Provided, by Type of Personnel[a]				
	Psychiatrist	Psychologist	Psychiatric Social Worker	Psychiatric Registered Nurse	Other
All nursing homes	26.1	14.8	19.3	4.6	19.9
Ownership					
Not for profit	22.1	18.9	22.0	1.9	12.9
For profit	26.3	13.2	17.9	4.3	22.2
Government	48.3	24.4	32.3	31.0	12.6
Bed capacity					
0–49	17.6	5.9	19.9	4.1	11.3
50–99	30.8	14.3	14.7	3.6	26.6
100–149	27.9	29.4	20.7	4.9	28.6
150 or more	45.2	24.1	24.8	8.5	22.9
Medicaid SNF certification					
No SNF beds	19.1	8.5	17.9	3.1	19.3
1–49% SNF beds	34.2	20.6	19.6	2.3	20.7
50–99% SNF beds	48.4	41.3	24.3	6.7	18.5
100% SNF beds	36.1	22.6	21.1	9.8	20.9
Facility chain status					
Part of chain	25.0	16.4	19.3	4.1	25.1
Independent nursing home	26.9	13.3	19.3	5.1	15.4

Note. SNF = skilled nursing facility.
Source. Compiled from data from the *National Medical Expenditure Survey, 1987: Institutional Population Component.*[2]
[a] Each figure represents solely the percentage of facilities in each category that reported services provided by that type of personnel; percentages do not total to 100% in any category or subcategory.

Note. Ind. = Independent.
Source. Compiled from data from the *National Medical Expenditure Survey, 1987: Institutional Population Component.*[2]

FIGURE 1—The provision of counseling or psychotherapy by nursing homes.

hospitals because they are unrepresentative in terms of staff and clients.[7,8] Finally, facilities for whom ownership could not be determined were excluded. The resident sample excluded all persons who were reported to be comatose. Thus, our estimates may differ slightly from those reported elsewhere.[4]

Results

The results depict the provision of mental health services and residents' use of services. Figure 1 shows the percentage of nursing homes providing counseling or psychotherapy and the proportions of contract versus other arrangements both for the total sample and according to ownership category. As can be seen, slightly fewer than 60% of the nation's nursing homes, as estimated from the National Medical Expenditure Survey sampling frame, provided some counseling or psychotherapy service, with a plurality of homes (26.5%) providing services via contract with an outside therapist. There were clear differences according to ownership category; government facilities were most likely to provide counseling (82.7%).

Table 1 reports on nursing homes' provision of mental health services. Larger nursing homes were more likely to provide some counseling services, although the percentage of nursing homes providing counseling services among nursing homes with 150 or more beds actually was lower than it was among nursing homes with 100 to 149 beds.

Table 2 shows the types of personnel providing mental health services in nursing homes by ownership category, bed capacity, Medicaid skilled nursing facility certification, and chain status. More than

TABLE 3—Receipt of Mental Health Treatment by Nursing Home Residents

	Diagnostic Category							
	Dementia Only		Dementia Complicated		Other Disorders		No Disorder	
	No. Residents	% Receiving Treatment	No. Residents	% Receiving Treatment	No. Residents	% Receiving Treatment	No. Residents	% Receiving Treatment
All nursing homes	399 512	6.2	211 641	26.1	224 615	35.1	609 862	5.5
Ownership								
Not for profit	78 809	8.7	48 611	35.7	38 865	31.5	145 295	5.2
For profit	292 213	4.9	147 361	23.4	168 637	34.9	423 974	5.5
Government	28 490	12.2	15 669	22.2	17 113	45.0	40 593	7.1
Bed capacity								
3–49	33 153	4.4	18 334	25.1	28 402	53.3	71 053	8.6
50–99	103 847	4.0	52 353	24.8	70 018	35.4	138 657	2.5
100–149	124 485	5.1	69 482	20.0	60 637	20.7	209 482	6.2
150 or more	138 028	9.2	71 472	33.3	65 558	40.1	190 671	5.8
Medicaid SNF certification								
No SNF beds	142 834	4.3	79 172	20.4	99 308	43.2	256 146	5.7
1–49% SNF beds	64 475	3.1	27 822	26.5	33 899	25.7	101 515	3.2
50–99% SNF beds	60 556	6.9	33 997	53.9	24 521	35.6	66 344	4.7
100% SNF beds	131 644	9.4	70 651	18.9	66 887	27.5	185 857	6.9
Facility chain status								
Part of chain	207 582	4.8	110 088	19.7	120 401	29.9	343 576	5.9
Independent nursing home	191 930	7.8	101 553	33.1	104 214	41.0	266 286	5.0

Note. SNF = skilled nursing facility.
Source. Compiled from data from the *National Medical Expenditures Survey, 1987: Institutional Population Component*.[2]

a quarter of the facilities reported that counseling was provided by a psychiatrist. The second most frequently cited provider was other mental health professionals, a category generally considered to represent workers with a bachelor's degree of some type. In general, government facilities were the most frequent users of psychiatrists and psychologists, whereas for-profit facilities were the most frequent users of "other mental health workers."

The provision of counseling or psychotherapy also clearly differed according to bed capacity, Medicaid certification, and chain status. Smaller facilities tended to rely more heavily on psychiatric social workers, although nearly half of the facilities reporting the provision of counseling reported of the use of a psychiatrist.

Table 3 indicates the number and percentage of nursing home residents who received mental health treatment by diagnostic categories and facility characteristics.

Seventy-five percent of the 835 000 nursing home residents with a mental disorder resided at a facility that provided some form of counseling or psychotherapy (data not shown). However, only 19% of these residents were reported to have received treatment. The rates of treatment were low in all types of facilities, although rates in not-for-profit and government facilities were slightly higher than those in for-profit facilities.

A similar pattern emerged when the experience of nursing home residents with specific mental disorders was examined. Roughly 70% of the residents with dementia only lived in a facility that provided counseling. However, only 5% to 12% received treatment, with those in governmental facilities being more than twice as likely to receive services than those in for-profit settings.

Discussion

This article summarizes the most current national data on the provision and use of mental health services in nursing homes. Three findings from the study are most salient. First, although more than three quarters of the residents with a mental disorder resided at a nursing home that provided mental health services, fewer than one fifth of such residents actually received any mental health services from the nursing homes at any time during the year, according to a very lenient definition of service receipt. In contrast, recent estimates suggest that fewer than 5% of nursing home residents receive mental health services within a month.[1]

Second, these data suggest that the types of personnel used to provide counseling and psychotherapeutic services vary by facility ownership and facility size. Not-for-profit facilities tend to rely on psychiatrists, psychologists, and social workers, whereas government facilities employ the highest percentage of psychiatrists. For-profit facilities tend to use more providers with little advanced training in mental health care. Smaller facilities rely more heavily on psychiatric social workers.

Third, the data are consistent with earlier reports of a substantial level of unmet need for mental health care within the nursing home setting.[1,9] Almost 65% of persons with mental health problems uncomplicated by dementia received no treatment while residing in a nursing home.

Two limitations of this study should be noted. First, these data were collected before the implementation of the Nursing Home Reform Act of 1987, with its emphasis on the provision of active treatment for mentally ill nursing home residents. It may be that higher rates of treatment will be revealed in the next National Medical Expenditures Survey, after implementation of the Omnibus Budget Reconciliation Act of 1987. At the very least, these data provide a benchmark for assessing the impact of the Omnibus Budget Reconciliation Act of 1987, as do earlier estimates.[1]

Second, there are reasonable concerns regarding the validity of the psychiatric diagnoses in the National Medical Expenditures Survey data. However, more than 90% of the staff-reported psychiatric diagnoses were confirmed by diagnoses in the residents' medical records. Thus, the validity of the findings is ensured to the extent that the medical records are accurate. However, there is a fiscal incentive for nursing homes to underdiagnose mental illness among their residents.[10,11] Thus, the data presented here may underestimate the rates of mental disorders and the effective provision of mental health services in nursing homes. □

Acknowledgment

Andrea Streit's work on this project was supported in part by grant 3-T32AG00048-13 to The Pennsylvania State University from the National Institute on Aging.

References

1. Burns BJ, Wagner R, Taube JE, Magaziner J, Permutt T, Landerman LR. Mental health service use by the elderly in nursing homes. *Am J Public Health.* 1993;83:331–337.
2. National Center for Health Services Research and Health Care Technology Assessment. *National Medical Expenditure Survey, 1987: Institutional Population Component* [public use tapes 6 and 7]. Rockville, Md: Westat, Inc [producer]; 1987. Ann Arbor, Mich: Inter-University Consortium for Political and Social Research [distributor]; 1990.
3. Cohen J, Dubay L. The effects of Medicaid reimbursement method and ownership on nursing home costs, case mix, and staffing. *Inquiry.* 1990;27:183–200.
4. Lair T, Lefkowitz D. Mental health and functional status of residents of nursing and personal care homes. In: *National Medical Expenditure Survey Research Findings 7.* Rockville, Md: Public Health Service, Agency for Health Care Policy and Research; 1990. DHHS publication PHS 90-3470.
5. Edward W, Edwards B. Questionnaires and data collection methods for the institutional population component. In: *National Medical Expenditure Survey Methods 1.* Rockville, Md: Public Health Service, National Center for Health Services Research and Health Care Technology Assessment; September 1990. DHHS publication PHS 90-3470.
6. Keeler EB, Kane R, Solomon DH. Short- and long-term residents of nursing homes. *Med Care.* 1981;14:363–369.
7. Burns BJ, Larson DB, Goldstrom ID, et al. Mental disorders among nursing home patients: preliminary findings from the national nursing home survey pretest. *Int J Geriatr Psychiatry.* 1988;3:27–35.
8. Smyer MA, Brannon D. The nursing home industry. In: Smyer MA, Cohn MD, Brannon D, eds. *Mental Health Consultation in Nursing Homes.* New York, NY: New York University Press; 1988:24–47.
9. Rovner B, German P, Broadhead J, et al. The prevalence and management of dementia and other psychiatric disorders in nursing homes. *Int Psychogeriatr.* 1990; 2(1):13–24.
10. Frisman LK, McGuire TG. The economics of long-term care for the mentally ill. *J Soc Issues.* 1989;45:119–130.
11. Swan JH, Gerard LE. Reimbursement and funding systems. In: Lurie EE, Swan JA, and associates, eds. *Serving the Mentally Ill Elderly.* Lexington, Mass: Lexington Books; 1987:139–161.

Psychotic Inpatients' Social Class and Their First Admission to State or Private Psychiatric Baltimore Hospitals

Carles Muntaner, MD, PhD, Paula Wolyniec, MA, John McGrath, MA, and Ann E. Pulver, ScD

ABSTRACT

Social class differences were investigated among patients admitted to public and private psychiatric hospitals. Participants included first admission White psychotic men admitted to Baltimore metropolitan area hospitals between 1983 and 1989. After adjusting for age and diagnosis, patients with low levels of skills/credentials were found to be more likely than patients with higher levels to be admitted to state psychiatric hospitals. These findings underscore the persistence of social class as a determinant of differences in the use of psychiatric care. (*Am J Public Health.* 1994;84:287–289)

Introduction

The study of the social class of the severely mentally ill in relation to their use of mental health services has a long tradition in public mental health research.[1–5] Clearly, social class is one of the determinants of health care usage. Through its effect on income,[6] for example, social class might determine the availability of health insurance.[7] Additionally, past studies of mental health hospital admissions have shown that social class has an influence on the type of facility to which mental patients are admitted.[e.g.,2] These studies have focused primarily on the most severe forms of mental disorder (e.g., psychotic disorders, schizophrenia in particular), which require long periods of hospitalization.[8,9] The aim of the present research was to investigate whether, among White male psychotic patients, social class is associated with first admission to state or private hospitals in the Baltimore metropolitan area.

Methods

Private, state, and community general hospitals with psychiatric units or service departments in the Baltimore metropolitan area were contacted for recruitment of potential participants.[10] Five of six state hospitals (with capacities of 48, 54, 350, 367, and 480 beds) and two of three private hospitals (with capacities of 220 and 322 beds) agreed to allow their patients to participate. The single private hospital that was not approached serviced only women.

Among Baltimore area community general hospitals with psychiatric units, only half were included in the original design. Because it was felt that valid inferences about patients from such hospitals could not be made if only half the facilities participated, patients recruited from these facilities were excluded from the present

Carles Muntaner is with the National Institute of Mental Health, Bethesda, Md. Dr Muntaner, Paula Wolyniec, John McGrath, and Ann E. Pulver are with the Epidemiology Genetics Program, Department of Psychiatry and Behavioral Sciences, The Johns Hopkins University School of Medicine, Baltimore, Md.

Requests for reprints should be sent to Carles Muntaner, MD, PhD, National Institute of Mental Health, Room B1-A 12, Federal Building, 7550 Wisconsin Ave, Bethesda, MD 20892.

This paper was accepted May 19, 1993.

TABLE 1—Distribution of White Male Psychotic Inpatients' First Admissions to State and Private Mental Hospitals, by Clinical and Sociodemographic Characteristics, Baltimore Area, 1983 through 1989

	State (n = 185)	Private (n = 63)
	No. (%)	No. (%)
Age[a], y		
16–35	164 (88.65)	49 (77.78)
36–55	21 (11.35)	13 (20.63)
≥56	0 (00.00)	1 (1.59)
Diagnosis[b]		
Other psychotic disorders	127 (68.65)	51 (80.95)
Schizophrenia	58 (31.35)	12 (19.05)
Family social class		
Uncredentialed	49 (26.49)	20 (31.75)
Marginal	63 (34.05)	21 (33.33)
Expert	22 (11.89)	16 (25.40)
Patient social class[c]		
Uncredentialed	80 (43.24)	20 (31.75)
Marginal	47 (25.41)	20 (31.75)
Expert	6 (3.24)	21 (33.33)

[a]Mantel-Haenszel chi-square test for trend for age $\chi^2 = 5.50$, $P = .01$.
[b]A best estimate research diagnosis was formulated by a psychiatrist 6 months after the index hospitalization and was based on multiple sources of information (modified Diagnostic Interview Schedule administered by a master's level clinical psychologist; medical record reviews; informant report). Chi-square test for diagnosis, $\chi^2 = 3.51$, $P = .06$.
[c]Mantel-Haenszel chi-square test for patient's social class $\chi^2 = 14.73$, $P = .001$.

TABLE 2—Odds Ratios (ORs) and 95% Confidence Intervals (CIs) for Clinical and Sociodemographic Factors of White Male Psychotic Inpatients' First Admission to a State Mental Hospital, Baltimore Area, 1983 through 1989[a]

	Unadjusted Estimates			Adjusted Estimates (Main Effects Model[b])		
	OR	95% CI	P	OR	95% CI	P
Age (in years)	0.98	0.94, 1.01	.23	0.98	0.94, 1.02	.39
Family social class						
Expert	1.00			1.00		
Marginal/uncredentialed	2.52	1.23, 5.19	.01	2.18	1.02, 4.68	.04
Patient social class						
Expert	1.00			1.00		
Marginal/uncredentialed	6.31	2.23, 17.88	<.001	5.23	1.77, 15.50	<.01
Diagnosis						
Other psychotic disorders	1.00[c]			1.00		
Schizophrenia	1.94	0.96, 3.91	.06	1.93	0.93, 3.99	.07

[a]Results from logistic regression analysis.
[b]Model chi-square (*df*): 20.56 (4).
[c]Indicates reference category.

analyses. Furthermore, because the sample did not include the private women's hospital, the analysis was restricted to men.

Research participants were recruited through systematic screening of all White inpatients with psychotic symptoms or diagnoses who were aged 16 and over and were first admitted to Baltimore metropolitan state and private mental hospitals (described above) between 1983 and 1989.[10] (The study was originally funded to test genetic hypotheses regarding psychotic disorders among White patients.) Two hundred forty-eight of these admissions were male.

Prior to their release from the hospital, consenting patients were administered a modified version of the Diagnostic Interview Schedule[11] by a master's level clinical psychologist; details of the diagnostic procedures have been reported elsewhere.[10] Of the eligible White male first admission patients, 62% of those from state hospitals and 55% of those from private hospitals consented to participate in the study. (Response rates including women were similar: 60% and 53% for state and private mental hospitals, respectively.) This 7% difference between state and private patients was not statistically significant ($z = 1.13$, $P = .26$). For each type of facility, respondents were similar to nonrespondents with respect to diagnosis and education, but among private hospital inpatients, respondents were more likely to be younger (mean age = 28 years) than nonrespondents (mean age = 33 years).

Interview questions on family and employment history asked for the patient's last occupation prior to admission, the patient's usual occupation, and the occupation of the head of household when the patient was 16. Last occupation and occupation of head of household were used to classify patients. Students were classified according to their last occupation. The patient's last occupation showed a high correspondence to his usual occupation, as found in previous studies.[12]

Because continuous measures of education disregard the significance of credentials,[13] we used Wright's skills/credentials measure of social class.[14,15] This measure yields three social class categories (expert, marginal, and uncredentialed), based on the respondent's occupation and educational credentials.[14,15] Experts are defined as those holding jobs that require scarce skills, particularly credentialed skills, which are rare relative to their demand by the market (e.g., lawyers). Experts enjoy a credential rent: their wages are usually above the cost of reproduction of their training. They would be worse off with an equal endowment of society's pool of skills and credentials. Marginal and uncredentialed class members are defined as those holding jobs that only require common skills, particularly uncredentialed skills (e.g., construction laborers).[16]

We used logistic regression to estimate the association between social class and type of psychiatric hospital to which the patient was admitted. Age was considered a potential confounder because it has been found to be associated both with social class[16] and with type of hospital to which psychotic patients are admitted.[8] Psychiatric diagnosis was treated as a two-category variable: either schizophrenia or other cases with psychotic symptoms and diagnoses. Included in the latter category were patients with the following diagnoses: bipolar, schizoaffective, schizophreniform, and paranoid disorders; major depression; atypical, brief reactive psychoses; organic and substance-induced psychoses; and borderline or schizotypal personality disorders. Because there were no differences between marginal and uncredentialed classes in preliminary logistic regression analyses, the two categories were collapsed for the final analyses.

Results

The distributions of clinical and sociodemographic variables (diagnosis, age, family and patient's social class) are presented in Table 1. Patients diagnosed with schizophrenia were more likely to be admitted to state mental hospitals than pa-

TABLE 3—The Interaction Between Patient's Social Class and Type of Psychotic Disorder for White Male Psychotic Inpatients' First Admission to a State Mental Hospital, Baltimore Area, 1983 through 1989[a]

Independent Variables[b]	Adjusted OR	95% CI	P
Expert social class			
Other psychotic disorder[c]	1.00		
Schizophrenia	1.07	0.22, 5.14	.93
Marginal/uncredentialed social class			
Other psychotic disorders	1.76	0.55, 5.63	.34
Schizophrenia	4.15	1.05, 16.41	.04

Note. Model chi-square (*df*): 13.52 (5).
[a]Results from multiple logistic regression analysis.
[b]Adjusted for age and family social class.
[c]Reference category.

tients with other psychotic disorders. Similarly, patients located in the marginal and uncredentialed social classes were more likely to be admitted to state than to private hospitals; the reverse was observed for patients in the expert social class.

Table 2 presents the unadjusted estimates of the odds of being admitted to a state hospital versus a private hospital for each independent variable in the study. Patients diagnosed with schizophrenia were almost twice as likely to have been admitted to a state psychiatric hospital than to a private one, although the estimate was not statistically significant. Patients whose families were from marginal or uncredentialed social classes were 2.5 times as likely and patients whose own social class at the time of admission was marginal or uncredentialed were more than six times as likely to be admitted to a state psychiatric hospital.

Table 2 also summarizes the results from the multiple logistic regression model. Odds ratios were similar to bivariate analyses, with no evidence of confounding.

Table 3 presents the results from our exploratory analysis of the interaction between social class and diagnosis. Adding the interaction term for patients' social class and diagnosis improved the fit of the model ($\chi^2_{(1)} = 7.04$, $P < .01$). Patients in the marginal/uncredentialed social class who suffered from schizophrenia were comparatively more likely to be admitted to state mental hospitals than to private hospitals.

Discussion

Our results reveal that social class strongly influences location of the first admission to a psychiatric hospital among White male psychotic patients. Furthermore, the greater likelihood of White male schizophrenics being admitted to state mental hospitals than to private hospitals varies by social class.

It is important to note that our inferences are restricted to White male psychotic patients in the Baltimore metropolitan area for the period 1983 to 1989. The fact that our population consisted exclusively of White men allowed us to isolate the effect of social class. However, this precluded the study of possible interactions between class, gender, and race[17,18] in determining the site of first hospital admission. Further research on the effects of social class on mental health admissions might explore the role of additional dimensions such as organizational and productive assets[16] in predicting the type of facility to which patients are admitted. Studies including general hospitals, multiservice mental health organizations, and residential treatment centers[9] could reinforce our conclusion that social class remains an important determinant of location of care among psychotic patients. □

Acknowledgments

This study was supported by National Institute of Mental Health grant MH35712.

We thank Elliot Liebow and Patricia Ocampo for their comments.

References

1. Faris REL, Dunham HW. *Mental Disorders in Urban Areas: An Ecological Study of Schizophrenia and Other Psychoses.* New York, NY: Hafner Publishing Co; 1939.
2. Hollingshead AB, Redlich FC. *Social Class and Mental Illness. A Community Study.* New York, NY: Wiley; 1958.
3. Kohn ML. Social class and schizophrenia: a review and a reformulation. *Schizophr Bull.* 1973;7:60–79.
4. Weyerer S, Dilling H, Kohl R, et al. Social class and mental disorders. A study of the use of medical services. *Soc Psychiatry.* 1982;17:133–141.
5. Kiesler CA, Sibulkin AE, Morton TL, et al. Characteristics of discharges from nonfederal, short-term specialty hospitals and general hospitals with and without psychiatric and chemical dependency units: the hospital discharge survey data. *Health Serv Res.* 1991;25:881–906.
6. Wright EO. *Class Structure and Income Determination.* New York, NY: Academic Press; 1979.
7. Friedman E. The uninsured. From dilemma to crisis. *JAMA.* 1991;265:2491–2495.
8. Manderscheid RW, Sonnenschein MA, eds. *Mental Health USA, 1990.* Washington, DC: National Institute of Mental Health; 1990. DHHS publication ADM 90–1708.
9. Kiesler CA, Simpkins C. The de facto national system of psychiatric inpatient care. *Am Psychol.* 1991;46:579–584.
10. Pulver AE, Liang K-Y, Hendricks Brown C, et al. Risk factors in schizophrenia. Season of birth, gender and familial risk. *Br J Psychiatry.* 1992;160:65–71.
11. Robins LN, Helzer JE, Croughan J, et al. National Institute of Mental Health Diagnostic Interview Schedule. *Arch Gen Psychiatry.* 1981;38:381–389.
12. Illis WR, Swanson GM, Satariano ER, Schwartz AG. Summary measures of occupational history: a comparison of latest occupation and industry with usual occupation and industry. *Am J Public Health.* 1987;77:1532–1534.
13. Krieger N. Women and social class: a methodological study comparing individual, household, and census measures as predictors of Black/White differences in reproductive history. *J Epidemiol Community Health.* 1991;45:35–42.
14. Wright EO, Martin B. The transformation of the American class structure, 1960–1980. *Am J Sociol.* 1987;93:1–29.
15. Wright EO, Cho D. The relative permeability of class boundaries to cross-class friendships: a comparative study of the United States, Canada, Sweden and Norway. *Am Sociol Rev.* 1992;57:85–102.
16. Wright EO. *Classes.* London, England: Verso; 1985.
17. Williams DR, Takeuchi DT, Adair RK. Socioeconomic status and psychiatric disorder among Blacks and Whites. *Soc Forces.* 1992;71:179–194.
18. Mutchler JE, Burr JA. Racial differences in health and health service utilization in later life: the effect of SES. *J Health Soc Behav.* 1991;32:342–356.

Psychiatric and Substance Abuse Disorders among Male Urban Jail Detainees

Linda A. Teplin, PhD

ABSTRACT

This paper presents prevalence rates by race/ethnicity and age for nine psychiatric and substance use disorders found in a random sample of 728 male jail detainees. Two thirds of the sample detainees had had a disorder other than antisocial personality during their lifetimes; half of these had had an episode within 2 weeks of the interview. More than 30% currently had either a severe mental disorder or a substance use disorder. Detainees with severe mental disorders or substance use disorders were most often in jail because they had committed nonviolent crimes. Policy implications of the results are discussed. (*Am J Public Health*. 1994;84:290–293)

Introduction

By law, mentally ill inmates must be treated for their disorders,[1-3] ideally receiving care comparable to what they would receive in the community.[4] Yet, as the Government Accounting Office notes, without better epidemiologic data, we cannot know how best to use the correctional system's scarce mental health resources.[5] The meteoric growth of jail populations[6] makes accurate epidemiologic data increasingly important.

Despite this need, there are few reliable data. Prevalence estimates of psychiatric disorders in jails vary widely[7-18] because of methodological inconsistencies across studies.[9] Moreover, no methodologically sound study has compared detainees' disorders with the crimes that precipitated their arrest. Such data are critical to planning treatment strategies for mentally ill jail detainees because the severity of the alleged offense determines, in part, the detainee's pretrial psychiatric treatment.[19,20]

This is the second of two epidemiologic papers analyzing data collected in 1983 and 1984 on 728 male jail detainees. The first paper was based on a subsample (n = 627) that could be compared with general population data from the Epidemiologic Catchment Area program; the analysis omitted 101 subjects who were Hispanic, younger than 18 years, or older than 60 years. Only schizophrenia and the major affective disorders were examined; the rates of these disorders among jail detainees were two to three times higher than general population rates.[9] In this paper prevalence rates are presented for the entire sample (n = 728) and by race/ethnicity (White, Black, and Hispanic) and age for nine current (within 2 weeks) and lifetime psychiatric disorders. The subject's arrest charge is also examined to see whether detainees with psychiatric disorders are in jail for more serious crimes than detainees with no disorder.

Methods

The subjects were a randomly selected, stratified sample (approximately one half misdemeanants, one half felons) of 728 male detainees interviewed at intake at Cook County Department of Corrections in Chicago, Ill. Of 767 detainees approached, 35 (4.6%) refused to participate. Four other subjects were excluded: two "duplicate" subjects (they were rearrested and randomly selected again) and two others who appeared to be confabulating their responses. The final sample total was 728. The National Institute of Mental Health Diagnostic Interview Schedule, Version III, which was developed for the Epidemiologic Catchment Area program, was used.[21] Because the sample was stratified by charge, all reported parameter and variance estimates were weighted to reflect the jail's true misdemeanant-to-felon ratio (55.9% misdemeanants and 44.1% felons). All reported standard errors and tests of significance have been corrected for these design effects.[22] Additional methodological information is available in the previous paper.[9]

Results

Nearly two thirds of the sample currently had one of the disorders listed in Table 1, and nearly 35% had a current disorder other than antisocial personality disorder. Rates of substance use disorders were extremely high: 29.1% (current) and 61.3% (lifetime). In the total sample, the proportions with either a severe mental disorder or a substance use disorder were 30.7% (current) and 61.8% (lifetime). Many subjects had both. Note that substance use *disorder* is more serious than *use*. Not surprisingly, antisocial personality disorder affected nearly 50% of the sample. As in the Epidemiologic Catchment Area study,[23] the rate of cognitive

The author is with the Psycho-legal Studies Program, Department of Psychiatry and Behavioral Sciences, Northwestern University Medical School, Chicago, Ill.

Requests for reprints should be sent to Linda A. Teplin, PhD, Psycho-legal Studies Program, 215 E Chicago Ave, Suite 708, Chicago, IL 60611-2610.

This paper was accepted September 30, 1993.

impairment found in the present study (2.0%) may have been inflated because of testing bias, so it was eliminated from the analysis. Given the legal mandate to treat currently ill detainees, the tables present only current disorders.

Table 2 shows the prevalence rates for three categories of race/ethnicity: White (non-Hispanic), Black (non-Hispanic), and Hispanic. Overall F tests were performed to see whether there were differences in prevalence rates among the three groups; *t* tests between groups were performed only when the overall F test was significant.[24] Table 2 shows some interesting racial/ethnic differences. For example, Whites had higher rates than Blacks for many disorders. Moreover, Hispanics had significantly higher rates than Blacks for schizophrenia and drug use disorder. Note that several large estimated differences in Table 2 are not significant, possibly because the small n's within categories cause standard errors to be quite large.

Table 3 presents the prevalence rates by age group. To test for age differences, an overall F test of significance was conducted; when differences were significant, F tests for nonlinearity were performed.[25] Again, for some age groups, the standard errors for the specific severe disorders are quite large relative to the estimated rates. Overall, the prevalence of many disorders increased with age.

Table 4 shows the bivariate relationship between current disorder (the unit of analysis) and the subject's most serious current charge: violent felony, nonviolent felony, violent misdemeanor, or nonviolent misdemeanor. Chi-square tests were used to compare the current charge distribution of each of the diagnostic groups with that of the control group, detainees with no disorders. Only one comparison was statistically significant: the drug use disorder group had fewer arrests for violent crimes (misdemeanors and felonies) than did detainees with no disorder ($P < .01$). These data were also analyzed with the subject as the unit of analysis; each subject was categorized by his most severe diagnosis. The results (available from author) were the same.

TABLE 1—Current and Lifetime Prevalence Rates of Disorders among Male Jail Detainees (n = 728)

	Current[a]			Lifetime		
	%	No.	SE	%	No.	SE
Any DIS disorder[b]	62.40	429	1.65	76.93	553	1.36
Any except antisocial personality	34.96	247	1.64	67.47	489	1.71
Any severe disorder (schizophrenia, manic episode, major depressive episode) or substance use disorder	30.70	220	1.67	61.83	447	1.77
Severe cognitive impairment[c]	2.00	14	0.25	...	...	...
Severe disorder (schizophrenia, manic episode, or major depressive episode)	6.12	45	0.85	8.94	66	1.01
Schizophrenia/schizophreniform disorder	2.98	21	0.72	3.82	26	0.91
Manic episode	1.18	9	0.32	2.17	17	0.38
Major depressive episode	3.42	26	0.56	5.09	39	0.66
Dysthymic disorder[d]	...	...	...	8.53	59	1.24
Substance use disorder (drug or alcohol)	29.11	206	1.70	61.25	443	1.78
Drug abuse/dependence	15.26	116	1.14	32.37	236	1.73
Alcohol abuse/dependence	19.09	129	1.68	51.11	367	1.85
Anxiety/somatoform disorder[e]	11.62	86	1.10	21.02	156	1.42
Antisocial personality disorder	48.40	344	1.86	49.21	351	1.85

Note. Disorders are counted as present only if the subject met the "severe" or "definite" criteria of the Diagnostic Interview Schedule (DIS).[21] Possible or moderate cases are considered as no disorder. Subjects may meet criteria for more than one disorder; therefore numbers may sum to more than the total sample. Cell counts are based on unweighted data. Reported percentages are based on data weighted to reflect the actual ratio of misdemeanants to felons in the jail population. For this reason, the reported count divided by the total n does not equal the reported percentage. Reported percentages represent parameter estimates. Data are missing as follows: any current DIS disorder, missing for 26 subjects; any current disorder except antisocial personality, missing for 26 subjects; any current severe disorder (schizophrenia or major affective disorders) or substance abuse disorder, missing for 23 subjects; current anxiety/somatoform disorder, missing for 3 subjects; current obsessive-compulsive disorder, missing for 1 subject; current alcohol abuse/dependence, missing for 23 subjects; current phobic disorder, missing for 2 subjects.

[a]Within 2 weeks of interview.

[b]This category includes only those DIS disorders specifically reported in this table. Egodystonic homosexuality, sexual dysfunction, tobacco abuse, anorexia nervosa, pathological gambling, and transsexualism are excluded.

[c]This diagnosis is determined by a brief mental status examination. See text for discussion of the diagnostic validity of this disorder.

[d]Recency of dysthymic disorder is not assessed by DIS version 3.

[e]This category includes the following disorders: phobia, panic disorder, obsessive-compulsive disorder, and somatization.

Discussion

More than 30% of the sample had a current severe mental disorder or substance use disorder, and nearly 62% had had one of these disorders at some time during their lifetime. As in the general population, alcoholism was extremely prevalent.[23,26,27] It is unlikely that the prevalence rates were inflated by the stresses of incarceration because subjects were interviewed during jail intake. DSM-III-R diagnoses require the exhibition of symptoms for a minimum of 2 weeks (longer for some disorders) or the existence of a prodromal phase.[28]

The finding that Whites had higher rates of some disorders than Blacks suggests a possible bias in the criminal justice process. Blacks have higher arrest rates than Whites[29] and, because they are poorer,[30] may have difficulty making bail. Blacks, mentally ill or not, are more likely to be jailed than Whites. In contrast, perhaps only the most dysfunctional Whites cannot make bail and are jailed before trial.

In contrast to the Epidemiologic Catchment Area study,[31] the present study found extremely high rates of schizophrenia (9.46%) among Hispanics. Further research is necessary to determine whether this finding is valid or whether it is due to testing bias in this jail population.[32,33]

Overall, the youngest age group (16 through 21 years) had the lowest prevalence rates for almost every disorder. Most likely, this is because the youngest subjects had not yet lived long enough to develop the disorders. They can be expected to develop additional disorders as they age.

The majority of disordered subjects were arrested for nonviolent crimes. Many of the severely ill detainees could have been diverted into the mental health system before arrest, before trial, or after release.[19] The diversion options depend, in part, on the severity of the charge. The American Bar Association recommends that mentally ill persons accused of minor crimes (misdemeanors) be diverted to mental health facilities, not arrested.[20]

TABLE 2—Prevalence of Current Disorders among Male Jail Detainees (n = 728), by Race/Ethnicity

	White, Non-Hispanic (n = 87)		Black, Non-Hispanic (n = 588)		Hispanic (n = 47)		Overall F Test	Significance and Direction of Pairwise *t* Tests[a]
	%	SE	%	SE	%	SE		
Schizophrenia/schizophreniform disorder	3.50	2.3	2.46	.7	9.46	5.0	$P < .05$	Hispanic > Black ($P < .05$)
Manic episode	2.57	1.4	1.08	.3	. . .[b]			
Major depressive episode	9.56	2.7	2.87	.6	. . .[b]			White > Black ($P < .001$)
Dysthymic disorder	6.78	3.3	7.76	1.3	17.21	6.8	NS	
Substance use disorder (drug or alcohol)	48.59	5.6	25.71	1.8	37.80	2.3	$P < .0001$	White > Black ($P < .0001$)
Drug abuse/dependence	24.91	4.1	12.90	1.2	27.86	5.8	$P < .001$	White > Black ($P < .01$) Hispanic > Black ($P < .01$)
Alcohol abuse/dependence	33.43	6.0	17.34	1.8	17.10	6.6	$P < .01$	White > Black ($P < .01$)
Anxiety/somatoform disorder	17.27	4.8	10.83	1.5	19.21	6.9	NS	
Antisocial personality disorder	65.81	3.1	46.80	1.2	40.74	4.5	$P < .01$	White > Black ($P < .01$) White > Hispanic ($P < .01$)

Note. Column n's are based on unweighted data. Reported percentages are based on data weighted to reflect the actual ratio of misdemeanants to felons in the jail population. Reported percentages represent parameter estimates. Variances and standard errors are adjusted for design effects. Six subjects coded "Other" on race are excluded. Subjects may meet criteria for more than one disorder. See notes to Table 1 for details on other missing data. NS = not significant at $P < .05$.

[a]Where the overall *t* test was significant at $P < .05$, *t* tests were performed for each pairwise contrast.

[b]There were too few Hispanic subjects to calculate reliable rates for major affective disorder (manic episode and major depressive episode). No overall tests of significance for these disorders were done. Pairwise tests for the remaining White-Black contrast are reported.

TABLE 3—Prevalence of Current Disorders among Male Jail Detainees (n = 728), by Age

	Age Group									
	16–21 y (n = 232)		22–28 y (n = 268)		29–35 y (n = 183)		40+ y (n = 45)		F Test Probabilities	Test for Nonlinearity
	%	SE	%	SE	%	SE	%	SE		
Schizophrenia/schizophreniform disorder	1.97	1.1	3.58	1.3	2.65	1.4	. . .[a]		NS	
Manic episode	0.34	.3	0.55	.4	2.01	.8	. . .[a]		NS	
Major depressive episode	1.29	.6	5.09	1.1	4.44	1.3	. . .[a]		$P < .05$	NS
Dysthymic disorder	4.05	1.6	12.33	2.4	9.00	2.6	5.73	4.0	$P < .05$	$P < .05$
Substance use disorder (drug or alcohol)	19.78	2.7	26.33	2.7	37.87	3.6	56.24	7.0	$P < .001$	NS
Drug abuse/dependence	10.98	1.8	13.15	1.7	22.08	2.7	21.41	5.0	$P < .01$	NS
Alcohol abuse/dependence	11.24	2.4	19.10	2.7	22.35	3.5	43.88	8.0	$P < .001$	NS
Anxiety/somatoform disorder	13.52	2.1	11.03	1.7	9.98	2.1	12.28	4.3	NS	
Antisocial personality disorder[b]	38.75	3.3	52.31	3.0	54.61	3.7	47.13	7.2	$P < .01$	$P < .05$

Note. Column n's are based on unweighted data. Reported percentages are based on data weighted to reflect the actual ratio of misdemeanants to felons in the jail population. For this reason, the reported count divided by total numbers does not equal the reported percentage. Reported percentages represent parameter estimates. Subjects may meet criteria for more than one disorder. See notes to Table 1 for details on missing data. NS = not significant at $P < .05$.

[a]The number of subjects aged 40+ years was too small to generate reliable rates for any of the severe disorders. F tests for these disorders were performed on the three reported cells.

[b]The Diagnostic Interview Schedule does not assess antisocial personality disorder for persons younger than 18 years. This age group has been omitted from this analysis. The n for the 18–21-year-old age group is 176.

Persons charged with felonies (even nonviolent felonies) must be adjudicated. Nevertheless, many detainees accused of nonviolent felonies could be treated outside the jail after their pretrial hearing if appropriate programs were available in the community.

If current rates of growth (15% to 20% per year) continue,[34–36] the national jail census will exceed 1 million by 1995. The number of jail detainees who will require treatment in the future—particularly for alcohol and drug abuse—is staggering. This is a critical public health problem because jails are not equipped to provide treatment.[12] Many jail detainees have multiple disorders, further complicating their treatment.[37,38]

The jail is a crucial point of intervention for persons whose psychiatric and substance abuse problems may lead them to commit minor offenses.[7,8] Crisis intervention treatment within the jail may help break the recidivistic cycle. Despite the legal mandate to treat disordered jail detainees,[1–3] few jails have adequate resources to treat either the mentally ill[19] or substance abusers.[39] Drug and alcohol treatment services are urgently needed. A recent nationwide survey of 1737 jails showed that only 28% provided *any* drug treatment services, and only 7% provided a comprehensive level of treatment.[39]

The median jail stay is only 3 days.[40] Thus, although treatment programs are needed, jails must also improve linkages with community mental health and substance abuse programs. Unfortunately, existing services are insufficient, especially for substance abuse.[41] Comprehensive drug and alcohol treatment services for jail detainees after release must become a priority. Diversion may also help reduce our burgeoning jail population.

Public health administrators must en-

TABLE 4—Male Jail Detainees (n = 726) with a Current Arrest Charge (%), by Current Diagnosis

Current Diagnosis	Felony		Misdemeanor		Chi-Square Test[a]
	Violent	Nonviolent	Violent	Nonviolent	
Schizophrenia/manic episode (n = 26)	15.4	46.2	3.8	34.6	NS
Major depressive episode (n = 26)	19.2	46.2	15.4	19.2	NS
Both drug and alcohol abuse/dependence (n = 32)	12.5	43.8	9.4	34.4	NS
Drug abuse/dependence (n = 116)	10.3	55.2	4.3	30.2	$P < .01$
Alcohol abuse/dependence (n = 128)	14.1	36.7	11.7	37.5	NS
No disorder (n = 504)	18.5	39.0	9.0	33.5	
Total	17.1	41.0	8.5	33.3	

Note. Row n's are based on unweighted data. Totals may not equal 100% because of rounding. Two subjects with cognitive impairment were excluded. Subjects may meet criteria for more than one disorder. NS = not significant at $P < .05$.
[a]Chi-square tests for each disorder contrasted with the no-disorder group ($df = 3$).

sure that jail treatment programs meet recently developed mental health standards.[42] Jails will then be in compliance with the courts, and jail detainees will receive the most effective and humane treatment possible within the correctional setting as well as after release. □

Acknowledgments

This work was supported by Public Health Service grant RO1MH37988.

I would like to thank Dr Gary McCleliand for providing extensive advice and assistance in preparing and analyzing these data. Dr Karen M. Abram provided insightful comments on earlier versions of this paper. Laura Coats was invaluable in the editing and preparation of the manuscript.

References

1. *Bowring v Godwin*, 551 F2d 44 (4th Cir 1977).
2. *Estelle v Gamble*, 429 US 97 (1976).
3. *Jones v Wittenberg*, 323 F Supp 793 (ND Ohio 1971).
4. *Psychiatric Services in Jails and Prisons.* Washington, DC: American Psychiatric Association; 1989. Task Force Report 29.
5. *Mentally Ill Inmates: Better Data Would Help Determine Protection and Advocacy Needs.* Washington, DC: US Government Accounting Office; 1991. Publication GAO/GGD-91-35.
6. *Our Crowded Jails: A National Plight.* Washington, DC: US Dept of Justice, Bureau of Justice Statistics; 1988. Publication NCJ-111846.
7. Teplin LA. The criminalization of the mentally ill: speculation in search of data. *Psychol Bull.* 1983;94:54–67.
8. Teplin LA. The criminalization hypothesis: myth, misnomer or management strategy. In: Shah S, Sales B, eds. *Law and Mental Health: Major Developments and Research Needs.* Rockville, Md: National Institute of Mental Health; 1991.
9. Teplin LA. The prevalence of severe mental disorder among male urban jail detainees: comparison with the Epidemiologic Catchment Area program. *Am J Public Health.* 1990;80:663–669.
10. Lamb R, Grant RW. The mentally ill in an urban county jail. *Arch Gen Psychiatry.* 1982;39:17–22.
11. Valdisseri EV, Carroll KR, Hartl AJ. A study of offenses committed by psychotic inmates in a county jail. *Hosp Community Psychiatry.* 1986;37:163–166.
12. Teplin LA. Detecting disorder: the treatment of mental illness among jail detainees. *J Consult Clin Psychol.* 1990;58:233–236.
13. Bolton A. *A Study of the Need for an Availability of Mental Health Services for Mentally Disordered Jail Inmates and Juveniles in Detention Facilities.* Boston, Mass: Arthur Bolton Associates; 1976.
14. Swank GE, Winer DE. Occurrence of psychiatric disorder in a county jail population. *Am J Psychiatry.* 1976;133:1331–1333.
15. Guy E, Platt JJ, Zwerling I, Bullock S. Mental health status of prisoners in an urban jail. *Criminal Justice Behav.* 1985;12:29–53.
16. Schuckit MA, Herrman G, Schuckit JJ. The importance of psychiatric illness in newly arrested prisoners. *J Nerv Ment Dis.* 1977;165:118–125.
17. Nielsen ED. Community mental health services in the community jail. *Community Ment Health J.* 1979;15:27–32.
18. Whitmer GE. From hospitals to jails: the fate of California's deinstitutionalized mentally ill. *Am J Orthopsychiatry.* 1980;50:65–75.
19. Steadman HJ, McCarty DW, Morrissey JP. *The Mentally Ill in Jail: Planning for Essential Services.* New York, NY: Guilford Press; 1989.
20. *ABA Criminal Justice Mental Health Standards.* Washington, DC: American Bar Association; 1989.
21. Robins LN, Helzer JE, Croughan J, Ratcliff K. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. *Arch Gen Psychiatry.* 1981;38:381–389.
22. Cochran WG. *Sampling Techniques.* 3rd ed. New York, NY: John Wiley & Sons; 1977.
23. Burnam MA, Hough RL, Escobar JI, et al. Six-month prevalence of specific psychiatric disorders among Mexican Americans and non-Hispanic Whites in Los Angeles. *Arch Gen Psychiatry.* 1987;44:687–694.
24. Darlington RB. *Regression and Linear Models.* New York, NY: McGraw-Hill Publishing Co; 1990.
25. Draper NR, Smith H. *Applied Regression Analysis.* 2nd ed. New York, NY: John Wiley & Sons; 1981:424–427.
26. Robins LN, Helzer JE, Weissman MM, et al. Lifetime prevalence of specific psychiatric disorders in three sites. *Arch Gen Psychiatry.* 1984;41:949–958.
27. Karno M, Hough RL, Burnam MA, et al. Lifetime prevalence of specific psychiatric disorders among Mexican Americans and non-Hispanic Whites in Los Angeles. *Arch Gen Psychiatry.* 1987;44:695–701.
28. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed., revised. Washington, DC: American Psychiatric Association; 1987.
29. *Crime in the United States.* Washington, DC: US Dept of Justice, Federal Bureau of Investigation; 1990.
30. *Statistical Abstract of the United States (104th Edition).* Washington, DC: US Bureau of the Census; 1983.
31. Robins LN, Regier DA. *Psychiatric Disorders in America.* New York, NY: The Free Press; 1991.
32. Malgady RG, Rogler LH, Costantino G. Ethnocultural and linguistic bias in mental health evaluation of Hispanics. *Am Psychologist.* 1987;42:228–234.
33. Padilla AM, Ruiz RA. *Latino Mental Health: A Review of the Literature.* Washington, DC: US Dept of Health, Education, and Welfare; 1973. DHEW publication HSM 73-9143.
34. *Census of Local Jails, 1988.* Vol. I. Washington, DC: US Dept of Justice, Bureau of Justice Statistics; 1991. Publication NCJ-127992.
35. *Profile of Jail Inmates, 1989.* Washington, DC: US Dept of Justice, Bureau of Justice Statistics; 1991. Publication NCJ-129097.
36. Camp GM, Camp C. *Corrections Yearbook 1989.* South Salem, NY: Criminal Justice Institute; 1990.
37. Abram KM. The problem of co-occurring disorders among jail detainees: antisocial disorder, alcoholism, drug abuse, and depression. *Law Human Behav.* 1990;14:333–345.
38. Abram KM, Teplin LA. Co-occurring disorders among mentally ill jail detainees: implications for public policy. *Am Psychol.* 1991;46:1036–1045.
39. Peters RH, May RL, Kearns WD. Drug treatment in jails: results of a nationwide survey. *J Criminal Justice.* 1992;20:283–295.
40. *Census of Local Jails 1988.* Washington, DC: US Dept of Justice, Bureau of Justice Statistics; 1990. Publication NCJ-121101.
41. Austin J. Getting tough with treatment. In: *The Compiler.* Chicago, Ill: Illinois Criminal Justice Information Authority; 1992.
42. Weinstein HC. Psychiatric services in jails and prisons: who cares? *Am J Psychiatry.* 1989;146:1094–1095.

Enhancing the Self-Report of Alcohol Consumption in the Community: Two Questionnaire Formats

Abby C. King, PhD

ABSTRACT

Two questionnaire formats for assessing alcohol consumption in a community sample were compared. Subjects completed the Semi-Quantitative Food Frequency Questionnaire and a questionnaire specifically targeting alcohol use. Across all alcoholic beverages, subjects reported lower consumption on the alcohol questionnaire than on the food frequency questionnaire. The results suggest that food frequency questionnaires may provide a better means for enhancing self-report of alcohol use than questionnaires that target solely alcohol intake. (*Am J Public Health.* 1994;84:294–296)

Introduction

Assessment of alcohol consumption typically involves targeting alcohol use directly or obtaining alcohol data as part of a more general dietary assessment.[1–4] When a specific alcohol questionnaire is used, respondents may be more likely to attend to the questions, and a more accurate report may be obtained.[5] In contrast, embedding alcohol intake questions in a general diet questionnaire may neutralize the negative feelings toward alcohol reporting potentially evoked in some population segments[6,7] and/or increase the number of associated cues that can enhance recall.[8] Little is known about reporting issues for the moderate drinking levels that are of greater relevance for public health research.[9]

Direct comparisons of the above two questionnaire methods have been rare.[10] The purpose of this study was to determine whether embedding alcohol questions in a general diet questionnaire would lead to increased reporting of alcohol use among a community-based sample. A food frequency questionnaire was chosen since such questionnaires are currently one of the primary methods for measuring dietary intake in public health research.[11]

Methods

Subjects

Subjects originated from a sample of 357 men and women who agreed to participate in a study evaluating the effects of increases in community-based physical activity on health outcomes.[12] Residents of Sunnyvale, Calif, between 50 and 65 years old were recruited through a random-digit dialing community telephone survey and citywide promotion.[13]

Procedures

Prior to randomization, subjects underwent two evaluation sessions scheduled 3 days apart. All subjects completed a semiquantitative food frequency questionnaire. In addition, the final 237 subjects entering the study completed an alcohol consumption questionnaire. These 237 individuals constitute the sample of interest. The order in which each subject completed the two questionnaires (i.e., during session 1 or 2) was determined randomly.

Measures

Food frequency questionnaire. The Semi-Quantitative Food Frequency Questionnaire was used to measure dietary intake over the previous year.[14] This questionnaire has proven to be reproducible and valid across a variety of nutrients, including alcohol.[14,15] Among its advantages are ease of administration across large samples and specification of drink size for different types of alcoholic beverages.[14]

Alcohol consumption questionnaire. The alcohol consumption questionnaire was composed of the same questions as the food frequency questionnaire with respect to the frequency of beer, red wine, white wine, and liquor consumed over the past year.

Statistical Analysis

Two categories were defined for each questionnaire: nondrinkers (those choosing the "never or less than once monthly" category) and drinkers (those choosing any of the other eight alcohol use categories). The McNemar symmetry chi-square statistic for paired comparisons[16] was used to compare the frequencies obtained across the two questionnaires for each alcoholic beverage separately and for all types of alcoholic beverages combined. Cohen's kappa was applied to measure agreement across the two questionnaires.[16]

Subjects' alcohol responses were then grouped into four categories and compared across the two questionnaires.

The author is with the Department of Health Research and Policy and the Department of Medicine, Stanford University School of Medicine, Palo Alto, Calif.

Requests for reprints should be sent to Abby C. King, PhD, 730 Welch Rd, Suite B, Stanford University School of Medicine, Palo Alto, CA 94304-1583.

This paper was accepted June 23, 1993.

The categories were nondrinker (never or less than one drink monthly), light drinker (one drink per week through one drink per month), moderate drinker (one drink per day through two drinks per week), and heavy drinker (two drinks per day or more). These categories reflect the somewhat lower alcohol consumption rates typically reported by people in this older age group.[17]

Results

Descriptive data for the 237 subjects are shown in Table 1. Subjects were comparable on these variables to the larger study sample from which they were drawn, and they were similar to the northern California population from which they were recruited (Table 1).[12] The mean number of drinks per previous 2-week period reported by those indicating alcohol intake over the past year on the alcohol consumption questionnaire was similar to that of national samples in this age range.[18]

TABLE 1—Baseline Descriptive Statistics of Study Sample and a Representative Sample of the Target Community

	Study Sample		Community Sample[a]	
	Men	Women	Men	Women
No.	135	102	765	1109
Age, y (mean ± SD)	56.4 ± 4.1	57.0 ± 4.3	56.8 ± 4.4	56.9 ± 4.5
Education, y (mean ± SD)	15.8 ± 2.5	14.1 ± 2.4	15.2 ± 3.0	13.7 ± 2.7
Caucasian, %	89.1	88.7	88.4	92.2
Body mass index, kg/m^2 (mean ± SD)	27.2 ± 4.2	26.1 ± 5.1	26.0 ± 3.6	25.0 ± 3.6
Cigarette smokers, %	17.3	23.5	18.5	25.6
Alcohol intake over previous 2-week period, No. of drinks[b] (mean ± SD)	10.5 ± 8.6	6.3 ± 5.3	12.8 ± 16.4	6.5 ± 10.9

Note. There were no statistically significant differences between the two groups on the variables evaluated.
[a]Based on a random-digit dialing telephone survey of people 50–65 years of age residing in Sunnyvale, Calif, during the period in which the clinical trial recruitment was initiated.
[b]For subjects reporting some alcohol intake during the past year on the alcohol consumption questionnaire (the study sample included 83 women and 120 men; the community sample included 914 women and 658 men).

Classification of Drinkers and Nondrinkers on the Two Questionnaires

The amount of interquestionnaire agreement for the drinking/nondrinking classification is shown in Table 2. For all four types of alcoholic beverage, the alcohol consumption questionnaire resulted in significantly fewer individuals classified as drinkers than did the food frequency questionnaire (Ps < .05). Cohen's kappa ranged from .48 (white wine) to .70 (liquor), signifying fair to good agreement between the two questionnaires.[16] Comparable analyses conducted separately for each sex led to similar results.

When a nondrinker was defined as an individual having no drinks or less than one drink per month across all types of alcoholic beverages and a drinker was defined as someone having one drink or more per month, significant discrepancies between the two questionnaires continued ($P < .002$; $\kappa = .45$). While 83.5% of subjects were identified as drinkers on both questionnaires, there was discordance between the two questionnaires for 11% of the sample, with the alcohol consumption questionnaire identifying lower numbers of drinkers than the food frequency questionnaire.

Analyses indicated similar reporting patterns regardless of whether the food frequency questionnaire was completed on day 1 or day 2.

Relative Frequency of Alcohol Intake

In addition to identifying fewer drinkers relative to the food frequency questionnaire, the alcohol consumption questionnaire generally resulted in a lower reported frequency of intake among drinkers across all three drinking frequency categories (light, moderate, and heavy) (see Table 3).

Discussion

Although it was completed within 3 days of the food frequency questionnaire and used identically worded items and time frame, the alcohol consumption questionnaire consistently resulted in the identification of a smaller percentage of drinkers and a lower reported frequency of alcohol intake. Although clearly not representative of the US population as a whole in this age group, the study sample did provide a good representation of the subjects in the age range of interest residing in the target community, which was

TABLE 2—Level of Correspondence between the Alcohol Consumption Questionnaire and the Semi-Quantitative Food Frequency Questionnaire (Column Percentages) for Subjects Categorized as Nondrinkers or Drinkers, by Type of Alcoholic Beverage

	Alcohol Consumption Questionnaire							
	Beer*		Liquor***		Red Wine**		White Wine**	
Semi-Quantitative Food Frequency Questionnaire	Nondrinker % Agreement (No.)	Drinker % Agreement (No.)	Nondrinker % Agreement (No.)	Drinker % Agreement (No.)	Nondrinker % Agreement (No.)	Drinker % Agreement (No.)	Nondrinker % Agreement (No.)	Drinker % Agreement (No.)
Nondrinker	67.8 (82)	3.5[a] (4)	78.9 (86)	9.4[a] (12)	71.5 (108)	18.6[a] (16)	50.0 (34)	7.1[a] (12)
Drinker	32.2[a] (39)	96.5 (112)	21.1[a] (23)	90.6 (116)	28.5[a] (43)	81.4 (70)	50.0[a] (34)	92.9 (157)
Total	100 (121)	100 (116)	100 (109)	100 (128)	100 (151)	100 (86)	100 (68)	100 (169)

[a]Subjects for whom the two questionnaires were discordant.
*$P < .0001$; **$P < .001$; ***$P < .05$.

TABLE 3—Percentage of Subjects Categorized as Nondrinkers or as Light, Moderate, or Heavy Drinkers on the Semi-Quantitative Food Frequency Questionnaire (SFFQ) and Alcohol Consumption Questionnaire (ACQ), by Type of Alcoholic Beverage

	Nondrinkers, %		Drinkers, %					
			Light		Moderate		Heavy	
	SFFQ	ACQ	SFFQ	ACQ	SFFQ	ACQ	SFFQ	ACQ
Beer	36	51	42	31	17	15	5	3
Red wine	52	64	32	26	13	9	3	1
White wine	19	29	46	37	28	27	7	7
Liquor	41	46	30	29	24	19	5	6

Note. Percentages are rounded off to the nearest integer. Spearman rank correlation coefficients for the two questionnaires were .78 for beer intake, .62 for red wine intake, .74 for white wine intake, and .79 for liquor intake.

largely White and well educated. The results suggest that use of alcohol items embedded in a food frequency questionnaire as opposed to a questionnaire specifically targeting alcohol intake may result in higher estimates of usual consumption in middle-aged and older adults. □

Acknowledgments

This study was supported by US Public Health Service grant AG-00440 from the National Institute on Aging (awarded to Dr King) and US Public Health Service grant HL-36272 from the National Heart, Lung, and Blood Institute (awarded to William L. Haskell, PhD).

Comments by Drs David G. Altman, Steven B. Lovett, Diane Tribble, and Marilyn Winkleby on earlier drafts of this manuscript are gratefully appreciated. I thank Ms Yongsook Kim for help with data collection, Helen C. Kraemer, PhD, for statistical guidance, and the citizens of Sunnyvale, Calif, for their interest and participation.

References

1. Midanik LT. Validity of self-reported alcohol use: a literature review and assessment. *Br J Addict.* 1988;83:1019–1030.
2. Ferrence RG, Truscott S, Whitehead PC. Drinking and the prevention of coronary heart disease: findings, issues and public health policy. *J Stud Alcohol.* 1986;47:394–408.
3. Flegal KM. Agreement between two dietary methods in the measurement of alcohol consumption. *J Stud Alcohol.* 1990;51:408–414.
4. Hellerstedt WL, Jeffery RW, Murray DM. The association between alcohol intake and adiposity in the general population. *Am J Epidemiol.* 1990;132:594–610.
5. Erber J, Herman TG, Botwinick J. Age differences in memory as a function of depth of processing. *Exp Aging Res.* 1980;6:341–348.
6. Poikolainen K. Underestimation of recalled alcohol intake in relation to actual consumption. *Br J Addict.* 1985;80:215–216.
7. Sudman S, Bradburn NM. *Asking Questions.* San Francisco, Calif: Jossey-Bass; 1982.
8. McEvoy CL, Nelson DL, Holley PE, Stelnicki GS. Implicit processing in the cued recall of young and old adults. *Psychol Aging.* 1992;7:401–408.
9. Giovannucci E, Colditz G, Stampfer MJ, et al. The assessment of alcohol consumption by a simple self-administered questionnaire. *Am J Epidemiol.* 1991;133:810–817.
10. Baghurst KI, McMichael AJ. Evaluation of questionnaire methods of measurement of alcohol consumption in young Australians. *Community Health Stud.* 1978;3:135–139.
11. Willett W. *Nutritional Epidemiology.* New York, NY: Oxford University Press; 1990.
12. King AC, Haskell WL, Taylor CB, Kraemer HC, DeBusk RF. Group- vs home-based exercise training in healthy older men and women: a community-based clinical trial. *JAMA.* 1991;266:1535–1542.
13. King AC, Harris RB, Haskell WL. Clinical trial recruitment: study- vs. subject-contact. *CVD Epidemiol Newsletter.* 1988;43:38. Abstract.
14. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol.* 1985;122:51–65.
15. Willett WC, Reynolds RD, Cottrell-Hoehner C, Sampson L, Browne ML. Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. *J Am Diet Assoc.* 1987;87:43–47.
16. Wilkinson L. *SYSTAT: The System for Statistics.* Evanston, Ill: SYSTAT Inc; 1989.
17. Schoenborn CA, Cohen BH. Trends in smoking, alcohol consumption, and other health practices among U.S. adults, 1977 and 1983. *Vital Health Stat.* 1986;118. DHHS publication PHS 86-1250.
18. Colliver J, Grigson B, Barbano H, Dufour M, Malin H. NHANES I Epidemiologic Followup Study: methodological issues and preliminary findings. In: Spiegler D, ed. *Alcohol Use among U.S. Ethnic Minorities.* Rockville, Md: National Institute on Alcohol Abuse and Alcoholism; 1985:411–423. Research Monograph 18.

The Frequency and Nature of Alcohol and Tobacco Advertising in Televised Sports, 1990 through 1992

ABSTRACT

This study examines the frequency and nature of alcohol and tobacco advertising in a random sample of 166 televised sports events representing 443.7 hours of network programming broadcast from fall 1990 through summer 1992. More commercials appear for alcohol products than for any other beverage. Beer commercials predominate and include images at odds with recommendations from former Surgeon General Koop. The audience is also exposed to alcohol and tobacco advertising through the appearances of stadium signs, other on-site promotions, and verbal or visual brief product sponsorships. Moderation messages and public service announcements are rare. (*Am J Public Health.* 1994;84:297–299)

Patricia A. Madden, MPH, and Joel W. Grube, PhD

Introduction

Alcohol and tobacco advertising is of interest to public health professionals because of its potential effects on consumers. In particular, concern has been expressed that this advertising may influence children.[1–3] Consistent with such concerns, most research has found small but significant correlations between exposure to alcohol and tobacco advertising and youth drinking and smoking attitudes and behavior.[1,4–8] Research has also shown that children's exposure to televised beer advertisements occurs primarily through sports programming.[8] Although tobacco commercials were banned from electronic media in 1971, nonstandard tobacco advertising (e.g., stadium signs and other visual and verbal references) broadcast to the television audience provides a vehicle for the tobacco industry to circumvent the federal ban. The present study documents the frequency and nature of alcohol and tobacco advertising in television sports programming. Advertisers' use of celebrity endorsers and images of activities that can be dangerous when paired with alcohol use is also examined.

Methods

Sample

A random sample of college and major professional sports events (football, baseball, and basketball) broadcast on network television from fall 1990 through summer 1992 was videotaped. All sampled events were broadcast on weekends before 5 PM, with the exception of "Monday Night Football." Beginning in summer 1991, the sample was expanded to include other sports such as golf, auto racing, bicycling, and Olympic coverage. Overall, 35 college sports events, 66 major professional sports events, and 65 other sports events, representing 443.7 hours of broadcasting, were studied (see Table 1).

Measures

All standard and nonstandard advertisements were assessed. Nonstandard advertisements were categorized as (1) stadium signs, (2) other on-site promotions (physically or audibly present to fans at the event), or (3) brief product sponsorships (visible or audible to the television viewing audience only). Other on-site promotions included logotypes on the sides of racing cars and giant blow-up beer cans; the typical format for a brief product sponsorship was as follows: "This game is being brought to you by X company." Each sports event was divided into 1-minute intervals, and each interval was coded for the presence of each type of nonstandard advertising. All occurrences of moderation messages or warnings about improper use of alcohol imbedded within advertisements were recorded.

Two hundred eight unique or distinct beer commercials were identified within the sample of sports programming. These unique commercials were coded for the presence of (1) celebrity endorsers, (2) cars or driving, and (3) water activities. The water activities category included water sports (e.g., kayaking, jet skiing), as well as scenes with people in boats or near water (e.g., beach, lake, swimming pool).

Eighty-three percent of the sports events in our sample were viewed independently by two coders. Intraclass correlation coefficients[9] for the reported variables ranged from .91 to 1.0. When disagreements arose, the average of the two coders' judgments was taken as the final count for analysis.

Results

Alcohol

Commercials. Alcohol is the most commonly advertised beverage on tele-

The authors are with the Prevention Research Center, Pacific Institute for Research and Evaluation, Berkeley, Calif.

Requests for reprints should be sent to Patricia A. Madden, MPH, Prevention Research Center, 2150 Shattuck Ave, Suite 900, Berkeley, CA 94704.

This paper was accepted October 13, 1993.

Editor's Note. See related annotation by Mosher (p 180) and article by Grube and Wallack (p 254) in this issue.

TABLE 1—Sample of Sports Programming: Fall 1990 through Summer 1992

Type of Sport	No. of Events	Total Sample Hours, No. (%)
College	35	89.7 (20.2)
Football	11	37.9 (8.5)
Basketball	24	51.8 (11.7)
Major professional	66	204.5 (46.1)
Football	27	84.3 (19.0)
Basketball	19	53.5 (12.1)
Baseball	20	66.8 (15.1)
Other	65	149.5 (33.7)
Golf	29	62.1 (14.0)
Miscellaneous[a]	36	87.4 (19.7)
Total	166	443.7 (100.0)

[a]Golf and miscellaneous sports events were sampled beginning in the summer of 1991 and are, therefore, included in only 5 of the 8 seasons analyzed.

TABLE 2—Hourly Rate of Alcohol Commercials (Standard Error) by Type of Sport: Fall 1990 through Summer 1992

Type of Sport	Beer	Wine	Wine Cooler	Total
College	1.09 (0.14)	0.02 (0.02)	0.01 (0.01)	1.12 (0.14)
Football	0.92 (0.21)	0.03 (0.02)	0.00 ...	0.95 (0.22)
Basketball	1.22 (0.19)	0.02 (0.02)	0.02 (0.02)	1.26 (0.19)
Major professional	2.16 (0.13)	0.09 (0.03)	0.04 (0.02)	2.29 (0.13)
Football	1.67 (0.19)	0.21 (0.05)	0.04 (0.03)	1.92 (0.22)
Basketball	2.97 (0.13)	0.00 ...	0.00 ...	2.97 (0.13)
Baseball	2.13 (0.24)	0.02 (0.01)	0.09 (0.04)	2.24 (0.28)
Other	0.71 (0.13)	0.03 (0.01)	0.02 (0.01)	0.76 (0.13)
Golf	0.18 (0.09)	0.03 (0.02)	0.02 (0.02)	0.23 (0.09)
Miscellaneous	1.09 (0.22)	0.02 (0.02)	0.02 (0.02)	1.13 (0.22)
Total	1.46 (0.11)	0.05 (0.01)	0.03 (0.01)	1.54 (0.11)

TABLE 3—Hourly Rate of Nonstandard Alcohol and Tobacco Advertising (Standard Error) by Type of Sport: Fall 1990 through Summer 1992

Type of Advertising	College Sports	Major Professional Sports[a]	Other Sports[b]	Total
Alcohol				
Stadium signs	0.06 (0.03)	2.76 (0.33)	2.21 (0.85)	2.03 (0.83)
Other on-site promotions	0.00 ...	0.01 (0.01)	1.24 (0.52)	0.42 (0.50)
Brief product sponsorships	0.30 (0.07)	1.24 (0.07)	0.61 (0.20)	0.84 (0.17)
Tobacco				
Stadium signs	0.01 (0.01)	1.83 (0.33)	1.03 (0.65)	1.19 (0.33)
Other on-site promotions	0.00 ...	0.02 (0.01)	0.59 (0.26)	0.21 (0.10)
Brief product sponsorships	0.00 ...	0.00 ...	0.16 (0.07)	0.06 (0.03)

[a]Major league football, baseball, and basketball.
[b]Auto racing, bicycling, bowling, diving, figure skating, golf, gymnastics, ice hockey, skiing, tennis, track and field, volleyball, yachting, and mixed event programs such as "Wide World of Sports" that include boxing, horse racing, and so forth.

vised sports programming. Six hundred eighty-five alcohol (beer, wine, wine cooler) commercials were broadcast in the 443.7 hours of sports programs. This figure represents more than three quarters (77%) of all beverage commercials. Nonalcoholic beer products accounted for 4% of all beverage commercials, while all other types of beverages combined (soft drinks, coffee, juice, etc.) accounted for 19%.

Table 2 compares the hourly rates of standard commercials for alcoholic beverages among types of sport programs. Just over 1.5 alcohol commercials, on average, were broadcast for each hour of sports programming. Beer commercials predominated wine and wine cooler commercials ($P < .05$ in both cases; Friedman test).

About 10% of the unique beer commercials were product oriented, simply showing images of the bottle and advertising the good taste and quality of the beer. Celebrity endorsers (often athletes or entertainers) appeared in 15% of the commercials. Approximately 16% contained images of cars or other motor vehicles, and 26% showed water activities. In all, 37% of the commercials contained either driving or water-related images.

Nonstandard advertising. Table 3 presents the hourly rate of appearances of stadium signs, other on-site promotions, and brief product sponsorships. Overall, such nonstandard advertising was common, occurring 3.3 times per hour. Stadium signs and brief product sponsorships were more prevalent in major professional sports events than in college or other sports events ($P < .05$; analysis of variance of ranks). Other on-site promotions were found predominantly in the "other sports" category (e.g., boxing and auto racing with alcohol insignia on the clothing of the boxers and drivers as well as advertising on the floor of the boxing ring and on the cars).

Nonalcoholic beer advertisements. Thirty-nine broadcasts of nonalcoholic beer commercials were found. Stadium signs, other on-site promotions, and brief product sponsorships for nonalcoholic beer appeared as well, primarily in golf tournaments.

Moderation messages. Of 685 broadcasts of alcohol commercials, only 3 contained moderation messages. In addition, 10 of the brief product sponsorships (about 3%) contained moderation messages (Budweiser's "Know when to say when" and "Friends know when to say when" slogans). Two segments produced in a "public service announcement" manner by a beer company addressed underage drinking and drinking and driving. The sponsor's name was featured, but no product was visible, and the messages focused exclusively on prevention.

Public service announcements. Ten broadcasts of alcohol-related public service announcements were found. These segments addressed drinking and driving, alcohol-free prom/graduation nights, and teenagers' negative behavior after intoxication.

Tobacco

The primary way in which tobacco is advertised in sports programming is through appearances of stadium signs (see Table 3). Major professional sports programming involves the highest frequency of appearances of signs per hour, followed by other sports programming ($P < .05$). Most brief product sponsorships (76%) and other on-site promotions (80%) appeared in the four auto racing events in the sample; such nonstandard advertisements

appeared in 219 of the 467 one-minute intervals (47%) during the broadcast of the four races. Overall, nonstandard tobacco advertising occurred 1.5 times per hour. No antismoking public service announcements were broadcast.

Discussion

This study provides a basis for monitoring trends in the frequency and nature of alcohol and tobacco advertising in televised sports programming. Results show that alcohol advertising is common. Beer commercials predominate and often feature celebrity endorsers and activities such as driving or being near water. These images or themes are at odds with former Surgeon General Koop's 1989 recommendations that alcohol advertising not portray activities that are dangerous when combined with drinking and that such advertising avoid celebrity endorsers, especially those who appeal to young people.[10] Viewers are also exposed to alcohol advertising via stadium signs, other on-site promotions, and brief product sponsorships. Moderation messages and public service announcements relating to alcohol are rarely broadcast. Thus, sport viewers receive primarily positive messages about drinking.

An outright ban of all types of alcohol advertising on television has been suggested. However, some public health researchers have cautioned against making the same mistake with alcohol as may have been made with tobacco.[11] Cigarette consumption actually increased after the 1971 ban, perhaps because of the discontinuation of equal time antismoking spots; the freeing of large sums of money, which the tobacco industry used for diversification; and the locking out of new competition through the elimination of access to the national audience.[12]

Short of an advertising ban, various restrictions have been suggested. Limiting alcohol advertising during times children are most likely to view (weekend mornings, early prime time, and times during which sports programming is broadcast) is one option.[8] Prohibiting images of automobiles, race tracks, speed boats, and other references to driving and speed in alcohol commercials has been suggested.[10,13] Former Surgeon General Novello appealed to the tobacco and alcohol industries to stop using cartoons or other images that appeal to children. Counteradvertising, such as that currently targeting smoking in California, and health warnings in electronic alcohol advertising also have been suggested. Finally, the presence of nonstandard advertising for tobacco products indicates a need to reexamine the definition of advertising currently used in enforcing the cigarette advertising ban. Cigarette stadium signs, other on-site promotions, and brief product sponsorships certainly expose viewers to cigarette advertising. The fact that such promotions appear on television is at odds with the intent of the 1971 ban. □

Acknowledgments

This study was supported by National Institute on Alcohol Abuse and Alcoholism Research Center grant AA06282 to the Prevention Research Center, Pacific Institute for Research and Evaluation.

An earlier version of this paper was presented at the 119th Annual Meeting of the American Public Health Association, October 1991, Atlanta, Ga.

We gratefully acknowledge Bettina Friese for her assistance with the preparation of this paper.

References

1. Wallack L, Cassady D, Grube J. *TV Beer Commercials and Children: Exposure, Attention, Beliefs, and Expectations about Drinking as an Adult.* Washington, DC: AAA Foundation for Traffic Safety; 1990.
2. Fischer P, Schwartz M, Richards J Jr, Goldstein A, Rojas T. Brand logo recognition by children aged 3 to 6 years: Mickey Mouse and Old Joe the camel. *JAMA.* 1991;266:3145–3148.
3. DiFranza JR, Richards JW Jr, Paulman PM, et al. RJR Nabisco's cartoon camel promotes camel cigarettes to children. *JAMA.* 1991;266:3149–3153.
4. Atkin C, Neuendorf K, McDermott S. The role of alcohol advertising in excessive and hazardous drinking. *J Drug Educ.* 1983;13: 313–325.
5. Atkin C, Block M. *Content and Effects of Alcohol Advertising.* Washington, DC: Bureau of Tobacco, Alcohol, and Firearms; 1981. Report PB-82-123142.
6. Aiken PP, Eadie DR, Leather DS, McNeill REJ, Scott AC. Television advertisements for alcoholic drinks do reinforce under-age drinking. *Br J Addict.* 1988;83:1399–1419.
7. Klitzner M, Gruenewald P, Bamberger E. Cigarette advertising and adolescent experimentation with smoking. *Br J Addict.* 1991;86:287–298.
8. Grube JW, Wallack L. Television beer advertising and drinking knowledge, beliefs, and intentions among schoolchildren. *Am J Public Health.* 1994;84:254–259.
9. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychological Rep.* 1966;19:3–11.
10. *Surgeon General's Workshop on Drunk Driving: Proceedings, Washington, DC, December 14–16, 1988.* Washington, DC: US Dept of Health and Human Services; 1989.
11. Warner K. Clearing the airwaves: the 1979 cigarette ad ban revisited. *Policy Analysis.* 1979;5:435–450.
12. Schuster C, Powell C. Comparison of cigarette and alcohol advertising controversies. *J Advertising.* 1987;16(2):26–33.
13. Postman N, Nystrom C, Strate L, Weingartner C. *Myths, Men and Beer: An Analysis of Beer Commercials on Broadcast Television, 1987.* Falls Church, Va: AAA Foundation for Traffic Safety; 1987.

The Illegal Sale of Cigarettes to US Minors: Estimates by State

ABSTRACT

Data available from recent national surveys on the cigarette consumption and purchasing practices of teenage smokers were used to generate state-specific estimates of the number of teenage smokers and cigarette sales to minors. In 1991, approximately 2.7 million teenage cigarette smokers consumed an average of 28.3 million cigarettes per day (516 million packs per year). An estimated 255 million packs of cigarettes were sold illegally to minors in 1991. To make cigarettes and other tobacco products less accessible to minors, policymakers should consider implementing various legislative and economic measures such as banning cigarette vending machines and raising tobacco excise taxes. (*Am J Public Health.* 1994;84:300–302)

K. Michael Cummings, PhD, MPH, Terry Pechacek, PhD, and Donald Shopland

Introduction

Cigarette smoking kills more than 400 000 Americans each year.[1] Most tobacco victims are over the age of 35[1]; however, the addiction that fosters tobacco use begins for a majority of smokers during adolescence.[2] In an effort to deter minors from initiating and sustaining tobacco use, most states have enacted laws that restrict the sale of tobacco to minors.[3] Despite the existence of these laws, evidence indicates that most teenage smokers buy their own cigarettes with little difficulty.[4–10]

In an effort to promote greater awareness of the problem of youth tobacco use and access, this study presents estimates of teenage smoking and the illegal sale of cigarettes to minors in each of the 50 states and the District of Columbia.

Methods and Materials

Data collected in the 1989 Teenage Attitudes and Practices Survey of 12- to 18-year-old youth, sponsored by the Centers for Disease Control, the National Cancer Institute, and the American Cancer Society, were used to generate state-specific estimates of the number of teenage smokers.[11] The survey is based on a representative national sample of 9965 young people. Since our main interest in this study was estimating the sale of cigarettes to teenagers, we restricted our definition of a smoker to anyone who reported having smoked in the last week. This definition provides a conservative estimate of teenage smoking but is more likely to capture those who regularly purchase their own cigarettes.[7–10]

Table 1 shows the age-specific prevalence estimates obtained from the Teenage Attitudes and Practices Survey. Teenagers who reported having smoked in the last week were asked to indicate the amount they smoked daily. Daily cigarette consumption among current smokers is also reported in Table 1.

Census figures from 1990 were used to estimate the number of teenagers in each age group by state. Multiplying the cigarette use prevalence figures shown in Table 1 by the number of teens in each age group yields an estimate of the number of smokers between the ages of 12 and 18 years in each state. Multiplying the average number of cigarettes consumed daily from Table 1 by the number of teenage smokers in each age group provides an estimate of the number of cigarettes consumed daily by teens in each state.

Data from a 1989 study on the cigarette purchasing habits of teenage smokers conducted by the National Automatic Merchandising Association[5] (a trade association for companies that sell cigarettes from vending machines) were used to estimate the percentage of cigarettes consumed by teens that represent sales to minors. In that study, the purchasing habits of 1015 teenage smokers interviewed in shopping malls in 20 US cities were assessed. Seventy-two percent of teenage smokers indicated that they obtain their cigarettes by purchasing them, while an additional 19% reported that they depend on friends to supply them. Since the friends of most teenagers are themselves minors, this study indicates that 91% (72% + 19%) of all cigarettes consumed by teens are sold to minors. These findings are consistent with results from other studies that have examined the cigarette purchasing habits of teenage smokers.[10,12]

State laws vary in regard to the minimum age for selling cigarettes to minors.[3] The estimated number of packs of cigarettes illegally sold to minors was calculated separately by state according to each state's minimum age law (see Table 2). Multiplying the number of packs of cigarettes consumed by teens by 0.91 provides an estimate of the number of packs sold illegally to minors each day. Sales and tax

K. Michael Cummings is with the Department of Cancer Control and Epidemiology, Roswell Park Cancer Institute, Buffalo, NY. Terry Pechacek is with the Department of Social and Preventive Medicine, School of Medicine, State University of New York at Buffalo. Donald Shopland is with the National Cancer Institute, Rockville, Md.

Requests for reprints should be sent to K. Michael Cummings, PhD, MPH, Department of Cancer Control and Epidemiology, Roswell Park Cancer Institute, Elm and Carlton Sts, Buffalo, NY 14263.

This paper was accepted May 25, 1993.

TABLE 1—Teenagers' Reported Cigarette Use in the Previous 7 Days and Average Daily Cigarette Consumption, by Age

Age, y	Teenagers Reporting Smoking during Preceding Week[a] %	95% Confidence Interval	No. of Cigarettes Smoked per Day[a] Mean	95% Confidence Interval
12	0.7	0.3, 1.1	8.42	3.37, 13.47
13	2.5	1.6, 3.4	4.50	2.84, 6.16
14	7.1	5.6, 8.6	6.50	5.06, 7.94
15	11.8	9.8, 13.4	8.28	7.01, 9.54
16	13.7	11.8, 15.6	9.49	8.31, 10.67
17	17.9	15.8, 20.0	10.88	9.84, 11.92
18	25.4	22.8, 28.0	12.29	11.19, 13.39

Note. The mean number of cigarettes smoked daily was computed only for those who responded to the telephone survey. Software for Survey Data Analysis[20] was used in calculating confidence intervals.
Source. Data are from the 1989 Teenage Attitudes and Practices Survey.[11]
[a]As of November 1, 1989.

data compiled by the Tobacco Institute[13] for 1991 were used to determine estimates of sales and tax revenue generated from the illegal sale of cigarettes to minors on a state-by-state basis.

Results

Table 2 shows the estimated number of smokers 12 to 18 years of age by state and the District of Columbia. Overall, we estimate that approximately 2.7 million teenagers smoked an average of 28.3 million cigarettes per day (516 million packs of cigarettes per year).

Table 2 also provides estimates of sales and tax revenue derived from the illegal sale of cigarettes to minors. In 1991, four states, Louisiana, Missouri, Montana, and New Mexico, did not have tobacco age sale restriction laws (since 1991, Louisiana and Missouri have passed laws banning the sale of tobacco to those under the age of 18) and thus did not contribute to the overall estimate for illegal sales to minors. Taking into account differences in state laws regarding minimum age restrictions for selling cigarettes to minors, we estimate that, in 1991, 255 million packs of cigarettes were sold illegally to minors.

TABLE 2—Estimated Number of Teenage Cigarette Smokers and Illegal Sales to Minors, by State and the District of Columbia in 1991

	Estimated No. of Cigarette Smokers 12 to 18 Years of Age	Minimum Age for Sale,[a] y	Estimated No. of Cigarette Packs Sold Illegally to Minors Annually, Millions	Sales Revenue, Thousands of Dollars	Tobacco Taxes, Thousands of Dollars	
					State	Federal
Alabama	51 203	19	8.65	15 737	1 427	1 729
Alaska	5 956	19	1.01	2 181	293	202
Arizona	39 441	18	3.87	6 936	697	775
Arkansas	29 714	18	2.99	5 497	657	597
California	299 131	18	29.60	62 535	10 276	5 872
Colorado	35 605	18	3.46	5 882	692	692
Connecticut	33 733	18	3.33	6 865	1 500	667
Delaware	7 214	17	0.41	796	100	83
District of Columbia	5 421	16	0.16	305	48	58
Florida	125 038	18	12.34	24 552	4 182	2 468
Georgia	80 346	17	4.75	7 894	571	951
Hawaii	11 734	18	1.12	2 409	527	224
Idaho	12 818	18	1.28	2 207	231	257
Illinois	130 110	18	12.86	24 555	3 857	2 571
Indiana	67 576	18	6.68	11 084	1 035	1 335
Iowa	31 297	18	3.08	5 850	1 108	616
Kansas	27 247	18	2.65	4 849	636	530
Kentucky	45 659	16	1.43	2 072	43	523
Louisiana	52 061	None	...	...	...	...
Maine	14 183	18	1.42	2 872	469	284
Maryland	50 065	18	4.93	8 285	789	986
Massachusetts	60 988	18	5.87	10 977	1 526	1 174
Michigan	111 568	18	11.06	20 577	2 766	2 213
Minnesota	47 187	18	4.63	9 812	1 990	926
Mississippi	34 377	18	3.46	5 886	623	692
Missouri	57 709	None	...	...	...	...
Montana	9 119	None	...	...	...	...
Nebraska	18 053	18	1.76	3 211	476	353
Nevada	10 751	18	1.08	2 133	377	215
New Hampshire	12 572	18	1.20	2 119	301	241
New Jersey	83 251	18	8.26	16 604	3 004	1 652
New Mexico	18 231	None	...	...	...	...
New York	191 104	18	18.71	39 096	7 295	3 741
North Carolina	78 074	17	4.46	6 875	223	893
North Dakota	7 425	18	0.74	1 363	215	148
Ohio	127 209	18	12.58	21 634	2 264	2 516
Oklahoma	37 351	18	3.69	6 719	849	738
Oregon	29 719	18	2.95	5 556	828	591
Pennsylvania	130 450	18	12.87	23 936	3 989	2 574
Rhode Island	10 409	18	0.99	1 994	369	199
South Carolina	43 567	18	4.32	6 821	302	863
South Dakota	8 070	18	0.79	1 398	184	160
Tennessee	58 684	18	5.83	9 682	758	1 166
Texas	202 871	18	20.22	38 021	8 292	4 045
Utah	23 117	19	3.86	6 452	1 024	773
Vermont	6 608	18	0.62	1 056	117	124
Virginia	67 452	18	6.54	10 392	163	1 307
Washington	50 798	18	4.98	10 000	1 692	995
West Virginia	22 954	18	2.32	3 711	394	464
Wisconsin	54 263	18	5.35	9 523	1 688	1 070
Wyoming	5 270	18	0.53	824	63	105
Total	2 774 813		255.69	479 738	71 181	51 095

[a]As of July 1, 1991.

Discussion

The results of this analysis show that US teenagers spent $962 million on cigarettes in 1991. Our estimate of cigarette expenditures by teens is probably conservative since we did not consider tobacco use by teenagers who smoke less fre-

quently than once per week. However, even using this conservative approach to computing cigarette expenditures by teenagers, we estimate that, in 1991, the cigarette industry derived approximately $190 million in profit from sales to minors ($94.8 million from illegal sales). Of course, the profit derived from cigarette sales to teenagers is tiny in comparison with the profits obtained from sales to adults. The financial incentive for the tobacco industry to persuade teenagers to smoke is not the short-term profit gained from selling cigarettes to them but the expected stream of profit to be gained over a smoker's lifetime.

The federal government and state governments also benefit financially from tobacco sales to teenagers. In 1991, we estimate that combined federal and state taxes collected on the sale of cigarettes to teens between the ages of 12 and 18 years exceeded $240 million ($122 million derived from illegal sales to minors). The federal and state governments should consider earmarking tax revenues collected from the illegal sale of cigarettes to minors to programs enforcing tobacco age restriction laws.

The public supports stronger limits on children's access to and use of tobacco products.[14–16] Evidence from several studies[9,17,18] indicates that when tobacco age restriction laws are enforced, tobacco sales to minors can be sharply reduced. Two recent studies[19,20] also provide evidence that active enforcement of age restriction laws on tobacco sales can lower the prevalence of smoking among teenagers. Unfortunately, active enforcement of tobacco age restrictions laws is the exception rather than the rule in most US communities.

Until recently, the federal government had taken no action to regulate the access of minors to tobacco. In July 1992, Congress passed an important amendment to the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) Reorganization Act designed to discourage tobacco use by young people. The Synar Amendment (named for Congressman Mike Synar, who offered this amendment to the ADAMHA bill) provides that, as a condition of receiving block grant funds for substance abuse prevention and treatment, states must enact and enforce a law prohibiting the sale or distribution of tobacco products to individuals under 18 years of age. Specifically, states are required to check retailer compliance with the tobacco age sale law by conducting random, unannounced inspections of locations at which tobacco products are sold and to submit a report to the secretary of health and human services describing their activities aimed at enforcing the law during the preceding year. If a state fails to demonstrate a reasonable effort to enforce the law preventing tobacco sales to minors, then it will lose between 10% and 40% (depending on the year) of its ADAMHA funds. ADAMHA block grant awards tentatively set for fiscal year 1993 range from $2.0 million for Wyoming to $201.2 million for California. Since no state wishes to risk losing substantial federal funding, the Synar Amendment provides a new impetus for state governments to enact meaningful legislation and enforce existing laws to curtail tobacco sales to minors.

Finally, policymakers should also consider increasing taxes on cigarettes as a strategy to reduce teenage smoking. In comparison with adults, teenagers have relatively little disposable income and are therefore more sensitive to increases in prices.[21] A recently completed Canadian study found that a 10% increase in the relative price of cigarettes (over and above inflation) resulted in a 14% decline in the prevalence of smoking by 15- through 19-year-old teenagers.[22] □

References

1. Shultz JM. Smoking-attributable mortality and years of potential life lost—United States, 1988. *MMWR.* 1991;40:62–63, 69–71.
2. Centers for Disease Control. Differences in the age of smoking initiation between Blacks and Whites—United States. *MMWR.* 1991;40:754–757.
3. Centers for Disease Control. State laws restricting minors' access to tobacco. *MMWR.* 1990;39:349–353.
4. Kirn TF. Laws ban minors' tobacco purchases, but enforcement is another matter. *JAMA.* 1987;257:3323–3324.
5. *Findings from the Study of Teenage Cigarette Smoking Purchasing Behavior.* Chicago, Ill: Response Research Inc; 1989.
6. Forster JL, Kunt-Inge U, Jeffery RW. Sources of cigarettes for tenth graders in two Minnesota cities. *Health Educ Res.* 1989;4:45–50.
7. Centers for Disease Control. Accessibility of cigarettes to youths aged 12–17 years—United States, 1989. *MMWR.* 1992;41:485–488.
8. *Youth Access to Cigarettes.* Washington, DC: US Dept of Health and Human Services; 1990.
9. *Youth Access to Tobacco.* Washington, DC: US Dept of Health and Human Services; 1992.
10. Cummings KM, Pechacek TF, Orlandi M, Lynn WR. Where teenagers get their cigarettes: a survey of the purchasing habits of 13 to 16 year olds in 12 U.S. communities. *Tobacco Control.* 1992;1:264–267.
11. Centers for Disease Control. Teenage cigarette smoking among youth—United States, 1989. *MMWR.* 1991;40:712–715.
12. Stenwick RS, Fish DG, Manfreda J, Gelskey D, Skuba A. Where Manitoba children obtain their cigarettes. *Can Med Assoc J.* 1987;137:405–408.
13. *The Tax Burden on Tobacco.* Washington, DC: Tobacco Institute; 1991;26.
14. Forster JL, McBride C, Jeffrey R, Schmid TL, Pirie PL. Support for restrictive tobacco policies among residents of selected Minnesota communities. *Am J Health Promotion.* 1991;6:99–104.
15. Centers for Disease Control. Public attitudes regarding limits on public smoking and regulation of tobacco sales and advertising—10 U.S. communities, 1989. *MMWR.* 1991;40:344–345, 351–353.
16. Marcus SE, Emont SL, Corcoran RD, Giovino GA, Pierce JP, Waller MN, Davis RM. Public attitudes about cigarette smoking: results from the 1990 Smoking Activity Volunteer Executed Survey. *Public Health Rep.* In press.
17. Feighary E, Altman DG, Shaffer G. The effects of combining education and enforcement to reduce tobacco sales to minors. *JAMA.* 1991;266:3168–3171.
18. Cummings KM, Coogan K. Organizing communities to prevent the sale of tobacco products to minors. *Int Q Community Health Educ.* 1992;13:77–86.
19. Jason LA, Ji PY, Anes MD, Birkhead SH. Active enforcement of cigarette control laws in the prevention of cigarette sales to minors. *JAMA.* 1991;266:3149–3153.
20. DiFranza JR, Carlson RR, Caisee RE Jr. Reducing youth access to tobacco. *Tobacco Control.* 1992;1:58. Letter.
21. *Reducing the Health Consequences of Smoking: 25 Years of Progress—A Report of the Surgeon General.* Washington, DC: US Dept of Health and Human Services; 1989. DHHS publication CDC 89-8411.
22. Ferrence RG, Garcia JM, Sykora K, Collishaw NE, Farinon L. The effects of pricing on cigarette use among teenagers and adults in Canada, 1980–1989. In: *Health-Oriented Policy Options on Tobacco Taxes in the 1992 Ontario Budget.* Toronto, Ontario, Canada: Council for a Tobacco-Free Ontario and the Nonsmoker's Rights Association; 1992.
23. Shah BV. *Software for Survey Data Analysis (SUDAAN) Version 5.30.* Research Triangle Park, NC: Research Triangle Institute; 1989.

A2623-1

American Journal of Public Health

March 1994, Vol. 84, No. 3 Established 1911

ED KIANKA
HEALTH REPORTER
CONGRESSIONAL INFORMATION SVC
4520 EAST-WEST HIGHWAY 9375
BETHESDA MD 20814

APHA
American Public Health Association

Featuring Environmental Health

Editorials and Annotations

346 Editorial: Building Ventilation and Symptoms: Where Do We Go from Here? *M. J. Mendell and L. Fine*

348 Annotation: How Do We Get Enough Folic Acid to Prevent Some Neural Tube Defects? *S. A. A. Beresford*

350 Annotation: HIV Risk and Injection Drug Users—Evidence for Behavioral Change *D. D. Chitwood*

Articles

VARIOUS TOPICS

351 The Effects of Specific Medical Conditions on the Functional Limitations of Elders in the Framingham Study *A. A. Guccione, D. T. Felson, J. J. Anderson, J. M. Anthony, Y. Zhang, P. W. F. Wilson, M. Kelly-Hayes, P. A. Wolf, B. E. Kreger, and W. B. Kannel*

359 US Dietary Patterns Associated with Fat Intake: The 1987 National Health Interview Survey *A. F. Subar, R. G. Ziegler, B. H. Patterson, G. Ursin, and B. Graubard*

367 The Geographic Variation of Cancer Incidence in Ontario *S. D. Walter, S. E. Birnie, L. D. Marrett, S. M. Taylor, D. Reynolds, J. Davies, J. Drake, and M. Hayes*

377 Using CD4 Counts to Evaluate the Stages and Epidemiology of HIV Infection in South Carolina Public Clinic Patients *S. Luby, J. Jones, and J. Horan*

382 The Trading of Sex for Drugs or Money and HIV Seropositivity among Female Intravenous Drug Users *J. Astemborski, D. Vlahov, D. Warren, L. Solomon, and K. E. Nelson*

388 Changes in HIV-Related Information Sources, Instruction, Knowledge, and Behaviors among US High School Students, 1989 and 1990 *D. Holtzman, R. Lowry, L. Kann, J. L. Collins, and L. J. Kolbe*

394 Tobacco Use by Black and White Adolescents: The Validity of Self-Reports *K. E. Bauman and S. E. Ennett*

399 Smokeless Tobacco Use and Increased Cardiovascular Mortality among Swedish Construction Workers *G. Bolinder, L. Alfredsson, A. Englund, and U. de Faire*

405 Urinary Tract Infection during Pregnancy: Its Association with Maternal Morbidity and Perinatal Outcome *L. A. Schieve, A. Handler, R. Hershow, V. Persky, and F. Davis*

411 Maternal Attitude toward Pregnancy and the Risk of Neonatal Death *M. N. Bustan and A. L. Coker*

ENVIRONMENTAL HEALTH

415 A 6-Year Follow-Up of Behavior and Activity Disorders in the Taiwan Yu-cheng Children *Y.-C. J. Chen, M.-L. M. Yu, W. J. Rogan, B. C. Gladen, and C.-C. Hsu*

422 Air Recirculation and Sick Building Syndrome: A Blinded Crossover Trial *J. J. K. Jaakkola, P. Tuomaala, and O. Seppänen*

429 Nitrogen Dioxide Exposures inside Ice Skating Rinks *M. Brauer and J. D. Spengler*

434 The Respiratory Health Impact of a Large Urban Fire *M. Lipsett, K. Waller, D. Shusterman, S. Thollaug, and W. Brunner*

(continued on page 337)

Contents

American Journal of Public Health

March 1994, Vol. 84, No. 3 ISSN: 0090-0036 Established 1911

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

439 Dioxins and Dibenzofurans in Adipose Tissue of the General US Population and Selected Subpopulations *J. E. Orban, J. S. Stanley, J. G. Schwemberger, and J. C. Remmers*

446 Pesticide Exposures, Cholinesterase Depression, and Symptoms among North Carolina Migrant Farmworkers *S. Ciesielski, D. P. Loomis, S. R. Mims, and A. Auer*

Public Health Briefs

VARIOUS TOPICS

452 AIDS Risk Reduction and Reduced HIV Seroconversion among Injection Drug Users in Bangkok *D. C. Des Jarlais, K. Choopanya, S. Vanichseni, K. Plangsringarm, W. Sonchai, M. Carballo, P. Friedmann, and S. R. Friedman*

456 Racial Differences in Rural Adults' Attitudes toward Issues of Adolescent Sexuality *R. D. Horner, K. M. Kolasa, T. G. Irons, and K. Wilson*

459 The Weights and Heights of Mexican-American Adolescents: The Accuracy of Self-Reports *H. Davis and P. J. Gergen*

462 The Prevalence of Intrauterine Growth Retardation in Mexican Americans *H. Balcazar*

466 The Proportion of Veterans among Homeless Men *R. Rosenheck, L. Frisman, and A.-M. Chung*

469 The Driver's License List as a Population-Based Sampling Frame in Iowa *C. F. Lynch, N. Logsden-Sackett, S. L. Edwards, and K. P. Cantor*

ENVIRONMENTAL HEALTH

473 Neural Tube Defects in Jamaica following Hurricane Gilbert *E. M. W. Duff and E. S. Cooper*

476 The Use of Commercial Tanning Facilities by Suburban Minnesota Adolescents *J. A. Oliphant, J. L. Forster, and C. M. McBride*

479 Screening for Insecticide Overexposure under Field Conditions: A Reevaluation of the Tintometric Cholinesterase Kit *R. McConnell and R. Magnotti*

Public Health Then and Now

482 Editor's Comments: Plurality of Views *E. Fee*

482 The Strange Career of Managed Competition: From Military Failure to Medical Success? *H. Waitzkin*

490 Commentary: Setting the Record Straight—A Reply to Howard Waitzkin *A. C. Enthoven*

493 A Rejoinder from Waitzkin *H. Waitzkin*

Other Journal Departments

338 APHA Masthead

341 Advertisers' Index

342 What *AJPH* Authors Should Know

346 Journal Masthead

495 Letters to the Editor

502 A Thank-you to Reviewers

506 Book Corner

511 Association News: Policy Statements

523 Job Opportunities

American Public Health Association

American Journal of Public Health

American Public Health Association

The American Journal of Public Health, **published monthly, is the official Journal of the American Public Health Association, 1015 15th St, NW, Washington, DC 20005; (202) 789-5600; TDD (202) 789-5673. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.**

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*
Helen Rodriguez-Trias, MD, *Immediate Past President*
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Leonard M. Schuman, MD, *Vice-President (USA)*
John E. F. Hastings, MD, MPH, *Vice-President (Canada)*
Oscar Feo, MD, MSc, *Vice-President (Latin America)*
Jay H. Glasser, PhD, *Treasurer*
Allan G. Rosenfield, MD, *Chair, Executive Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*
Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*
Carol Easley Allan, PhD, RN (1997)
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Patricia P. Evans, MPH (1994)
Eugene Feingold, PhD, JD, *President*
Jay H. Glasser, PhD, *Treasurer*
Alan R. Hinman, MD, MPH (1995)
Margo D. Kerrigan, MPH (1995)
Joanne Lukomnik, MD (1996), *Vice-Chair*
Steven Uranga McKane, DMD, MPH (1995)
Carlos W. Molina, EdD (1997)
Linda Rae Murray, MD, MPH (1996)
Hila Richardson, DrPH, MPH (1996)
Helen Rodriguez-Trias, MD, *Immediate Past President*
Reed V. Tuckson, MD (1994)
Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*
Frances Kissling, *Chair, Action Board*
George E. Pickett, MD, MPH, *Speaker, Program Development Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*

The *American Journal of Public Health* (*AJPH*), published monthly, is the official Journal of the American Public Health Association, Inc, 1015 15th St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005. *AJPH* allotted $40; *The Nation's Health* allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters. Printed in the United States.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on the first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, *AJPH* Reprints, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the Journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The *American Journal of Public Health* is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

The Effects of Specific Medical Conditions on the Functional Limitations of Elders in the Framingham Study

Andrew A. Guccione, PhD, PT, David T. Felson, MD, MPH, Jennifer J. Anderson, PhD, John M. Anthony, BS, BA, Yuqing Zhang, MB, MPH, Peter W. F. Wilson, MD, Margaret Kelly-Hayes, EdD, RN, Philip A. Wolf, MD, Bernard E. Kreger, MD, MPH, and William B. Kannel, MD

ABSTRACT

Objectives. The purpose of this study was to identify associations between specific medical conditions in the elderly and limitations in functional tasks; to compare risks of disability across medical conditions, controlling for age, sex, and comorbidity; and to determine the proportion of disability attributable to each condition.

Methods. The subjects were 709 noninstitutionalized men and 1060 women of the Framingham Study cohort (mean age 73.7 ± 6.3 years). Ten medical conditions were identified for study: knee osteoarthritis, hip fracture, diabetes, stroke, heart disease, intermittent claudication, congestive heart failure, chronic obstructive pulmonary disease, depressive symptomatology, and cognitive impairment. Adjusted odds ratios were calculated for dependence on human assistance in seven functional activities.

Results. Stroke was significantly associated with functional limitations in all seven tasks; depressive symptomatology and hip fracture were associated with limitations in five tasks; and knee osteoarthritis, heart disease, congestive heart failure, and chronic obstructive pulmonary disease, were associated with limitations in four tasks each.

Conclusions. In general, stroke, depressive symptomatology, hip fracture, knee osteoarthritis, and heart disease account for more physical disability in noninstitutionalized elderly men and women than other diseases. (*Am J Public Health.* 1994;84:351–358)

Introduction

Medical comorbidity in the elderly is well documented.[1] Furthermore, an increase in the number of coexistent conditions among elders is directly associated with an increase in limitations in activities of daily living.[1] Associations between disease and physical disability in the elderly have been found for arthritis,[2–8] hip fracture,[9] low back pain,[8,10] diabetes,[6,7,11,12] hypertension,[8,12] coronary heart disease,[13,14] cardiovascular disease,[15] mild heart trouble,[6] heart attack,[11] shortness of breath,[6] peak expiratory flow rate,[16] leg cramps,[6] depression,[17,18] stroke,[6,7,11,19,20] visual impairment,[7] breast cancer,[21] and cognitive deficits.[22] There is, however, no reason to assume that all conditions have similar impacts on function.

Previously, the relationship between arthritis and physical functional limitations in the elderly has been shown to be limited to those activities that require the individual to use the afflicted joint.[2] Other conditions may also have task-specific impacts on function. For example, an individual with heart or lung disease may be limited in activities requiring endurance but not in those that require only dexterity. Thus, in a disabled elder with multiple chronic conditions, the inability to perform certain tasks may be related either to the broad effects of a single condition on most or all activities of daily living or to the independent effects of several conditions, each of which affects only a few activities. Although it is common to adjust for age, sex, and other sociodemographic variables when investigating the association between a particular medical condition and physical disability in the elderly, few investigators have attempted to sort out the independent impacts of specific diseases on the physical function of elders when these conditions coexist.[23] If one goal of medical care for the elderly is to control or diminish the effects of disease on physical functional status, then we must understand which specific medical conditions are associated with which specific functional limitations.

Many investigations of the relationship between disease and disability have relied on self-reported disease, potentially hindering accurate assessment of the connection between specific medical conditions and physical function. Previous investigations have also been limited by statistical approaches that count the number of comorbid conditions or the

Andrew A. Guccione, David T. Felson, Jennifer J. Anderson, John M. Anthony, and Yuqing Zhang are with the Arthritis Center, Boston University School of Medicine, Boston, Mass. Dr Guccione is also with Harvard Medical School and Massachusetts General Hospital, Boston. Dr Felson is also with the Boston City and University Hospitals, Boston. Peter W. F. Wilson is with the Framingham Heart Study. Margaret Kelly-Hayes and Philip A. Wolf are with the Department of Neurology, Boston University School of Medicine. Bernard E. Kreger is with the Framingham Heart Study and the Robert Dawson Evans Department of Clinical Research and Preventive Medicine, Boston University Medical Center. William Kannel is with the Section on Preventive Medicine, Boston University School of Public Health.

Requests for reprints should be sent to Andrew A. Guccione, PhD, PT, Physical Therapy Services, Massachusetts General Hospital, 15 Parkman St, WAC 128, Boston, MA 02114.

This paper was accepted June 23, 1993.

number of functional tasks affected, ignoring the potential task-specific nature of the relationship between disease and disability and the unequal effects of different conditions on function.

The overall aim of our investigation was to determine the independent effects of specific medical conditions on seven functional activities. Our study had three goals. The first was to identify the cross-sectional association in the Framingham cohort between specific medically validated diagnoses and functional limitations in two lower-extremity activities and five complex functional tasks. The second goal was to compare the risks of physical disability in elders for each medical condition, controlling for age, sex, and the presence of all other comorbid conditions. The final goal was to determine the proportion of disability in the cohort attributable to each medical condition, incorporating both the prevalence of the condition and the increased risk of functional limitation associated with it.

Method

Study Population

Our study population was the Framingham Study cohort. This cohort was enrolled between 1948 and 1951; the 5209 adults aged 28 through 62 constituted a one-third sample of Framingham residents in this age group. This group has been followed for more than 40 years; survivors are examined once every 2 years. With respect to mortality, fewer than 4% of the initial cohort of 5209 individuals have been lost to follow-up. At the time our data were collected, there were 2731 survivors. The data for this study are based on the 1826 elders who participated in the 18th biennial examination (1983 through 1985). The majority of nonparticipating survivors resided out of state and did not visit Framingham during the 2 years of the cohort examination.

Severe functional limitations in even the most basic activities of daily living have been documented for institutionalized elders.[24] We proposed to study higher levels of physical functional status, including instrumental activities of daily living such as grocery shopping and housekeeping, in noninstitutionalized elders. We limited our analyses to 1769 individuals who were not institutionalized at the time of their on-site participation in the 18th examination.

Definitions of Disease

All disease data were collected by trained study physicians and nurses and verified by a panel of physicians. We initially identified 20 medical conditions of potential interest in the data set available for study (Appendix A). We studied two musculoskeletal conditions: knee osteoarthritis and hip fracture. We allowed for two clinical presentations in our definition of knee osteoarthritis: (1) at least grade 2 radiographic changes according to the Kellgren and Lawrence criteria[25] with pain in or around the knee lasting at least a month in the year prior to examination; or (2) at least grade 3 radiographic changes (joint space narrowing plus osteophytes) with infrequent pain. In a previous study of knee osteoarthritis in the elders of the Framingham Study, it was found that those with infrequent symptoms and severe radiographic changes had increased odds of lower-extremity disability comparable to those of elders with symptoms but less severe radiographic changes.[2] Hip fracture was defined as fracture of the proximal femur prior to the 18th biennial examination.

There is evidence that obesity enhances the effects of arthritis on disability.[4] Therefore, we included obesity, defined as a body mass index (weight in kilograms divided by height in meters squared) of 30 or more, as the third medical condition in our preliminary analyses. Elders who used either insulin or oral hypoglycemics were counted as persons with diabetes.

Eight cardiovascular conditions were considered in the preliminary analysis of the relationship between a disease and disability. Hypertension was defined as a systolic blood pressure greater than or equal to 160 mm Hg and a diastolic blood pressure greater than or equal to 95 mm Hg at examination 18, without reference to treatment. Elders with cerebral vascular accidents, excluding transient ischemic attacks, prior to examination 18 were counted as persons with stroke. Specific criteria for angina pectoris, myocardial infarction on electrocardiogram, coronary insufficiency, and congestive heart failure have been developed for use at Framingham and are published elsewhere.[14] For the purposes of this study, we defined heart disease as any history of angina or myocardial infarction on electrocardiogram. We also included physician-diagnosed intermittent claudication as the final cardiovascular condition.

Three pulmonary disorders were defined for our preliminary analyses. First we examined dyspnea on exertion, although we recognized that it might be related to conditions other than lung diseases. Using spirometry performed at the previous biennial examination (1981 through 1983), we defined chronic obstructive pulmonary disease as ratio of forced expiratory volume in 1 second to forced vital capacity of less than 60%.[26] The third condition, symptomatic chronic obstructive pulmonary disease, was defined as chronic obstructive pulmonary disease on spirometry as defined above with dyspnea on exertion. Cancer was defined as any prevalent cancer except nonmelanoma skin cancer within the 5 years prior to an individual's visit at examination 18.

Mental status examinations are routinely conducted at Framingham.[27] As recommended by its developer, a score of less than 24 of a possible 30 on the Folstein Mini-Mental Status Examination was used to indicate cognitive impairment.[28] The 10-item Framingham screening measure of depressive symptoms is based on the CES-D (Center for Epidemiologic Studies–Depression Scale).[29] Each self-report item is scored dichotomously, indicating presence or absence of the symptom. We considered a score of 4 or higher at examination 18 as evidence of depressive symptomatology.

Data on cataracts and visual acuity were used to identify any association between visual impairment and functional limitations. We considered cataracts in either eye or both eyes. Although data on visual acuity at examination 18 were not available, we used data gathered at the 19th biennial exam on a subsample of elders and defined impairment as best-corrected vision worse than 20/40 in either eye.

Measures of Physical Disability

Data were collected on 16 measures of physical disability at examination 18. Six items measuring the most basic activities of daily living (e.g., bathing and toileting) were rejected for inclusion because we found little disability on these items, as might be anticipated in a sample of elders living independently in the community. Three items measuring

more complex activities (e.g., dialing a telephone) were rejected for the same reason. Of the seven remaining functional measures, two items covered activities explicitly involving the lower extremities: walking up and down stairs to the second floor and walking a mile. Each of these items was adapted from the Rosow-Breslau Functional Health Index[30] and had been used previously at Framingham. Five measures covered functional tasks requiring both upper and lower extremities: housekeeping; heavy home chores, including shoveling snow and washing windows; cooking; grocery shopping; and carrying bundles weighing 10 pounds. The item on heavy home chores came from the Rosow-Breslau Functional Health Index; the last item was taken from an instrument developed by Nagi.[31] The items on housekeeping, cooking, and grocery shopping had also been previously used at Framingham.[32]

For items from the Rosow-Breslau instrument, subjects were recorded as able or unable to do without help. The items on housekeeping, cooking, and grocery shopping asked whether the respondent could perform that activity entirely without help from another person. The item on carrying bundles was part of a physical performance battery administered by a study nurse.[33] For this item, subjects were scored as independent, dependent on an assistive device, dependent on human assistance, or unable to do. To make the definition of the dependent variable comparable across items, we chose the need for human assistance as the criterion of functional limitation for each of the seven dependent variables.

Data Analysis

To pare our list of diagnoses using both clinical and statistical criteria, we evaluated the strength of the association of each of the initial 20 conditions with disability in each of the seven functional tasks. We considered whether the condition had clinical relevance to any one of the functional activities (e.g., intermittent claudication and walking) and whether the age- and sex-adjusted relative risk was at least 1.5. On the basis of these criteria, we identified 10 prevalent medical conditions associated with physical disability in the Framingham cohort: knee osteoarthritis, hip fracture, diabetes, stroke, heart disease, congestive heart failure, intermittent claudication, chronic obstructive pulmonary disease, depressive symptomatology, and cognitive impairment. The specific criteria for these diagnoses are found in Appendix B.

We used logistic regression analyses to calculate prevalence odds ratios for dependence in performing each of the seven functional tasks. All elders were used as the referent group in these calculations so that these odds ratios are estimates of relative risk. An odds ratio was defined as the odds of functional dependence in performing a functional task among subjects with a specific medical condition divided by the odds of dependence in that task among all subjects. Because we had noted a high degree of association between heart disease and congestive heart failure, we did not adjust for the presence of heart disease in the sample when calculating the odds of disability for elders with congestive heart failure. We did, however, include both conditions when adjusting for all noncardiac conditions.

In each of the seven logistic regression equations for a particular condition, we controlled for sex, for age as a continuous variable, and for the presence of each of the 10 medical conditions to calculate an adjusted odds ratio (aOR). The adjusted odds ratio was computed with 95% confidence intervals based on maximum likelihood–derived standard errors.

To understand the impact of each condition in terms of its prevalence as well as its odds of disability, we also computed the proportion of dependence in a particular task attributable to each condition as an adjusted attributable fraction (AF).[34] We calculated the adjusted attributable fraction in this study using the formula

$$AF = [(p)(aOR - 1) \div aOR],$$

where p is the proportion of functionally limited elders with a specific disease.

We had obtained knee radiographs on 1416 subjects in our sample. To maintain the size of our sample for this study, we used age, sex, and knee symptoms to impute knee osteoarthritis to elders who had participated in examination 18 but did not have radiographs. Our data on osteoarthritis includes the imputation of knee osteoarthritis to 61 elders. We compared analyses with and without these imputed cases and found no differences; therefore, imputed cases of knee osteoarthritis are included in the results below.

TABLE 1—Age Breakdown of Study Sample (n = 1769)

	Men (n = 709)		Women (n = 1060)	
	No.	%	No.	%
≤69 y	260	36.7	353	33.4
70–79 y	328	46.3	518	48.9
≥80 y	121	17.1	188	17.7
Mean ± SD	73.4 ± 6.3		73.9 ± 6.4	
Range	64.1–92.1		63.8–94.6	

Results

The age breakdown of the sample is shown in Table 1. There were 1060 women and 709 men; the mean age was 73.7 years. The prevalence of the 10 medical conditions included in our analyses and the proportion of elders with at least one comorbidity for each disease are found in Table 2. We found heart disease and knee osteoarthritis the most prevalent and hip fracture and congestive heart failure the least prevalent of these diseases.

The prevalence of dependence in each of the seven functional activities is found in Table 3. Dependence was substantially more frequent for heavy home chores, which require upper and lower extremity use, and walking a mile, one of the two lower-extremity items; dependence in cooking was found least often.

The age-, sex-, and comorbidity-adjusted odds ratios for dependence in each activity are found in Table 4. Only stroke was significantly associated with dependence on human assistance for all seven tasks. Depressive symptomatology and hip fracture were each significantly associated with functional limitations in five tasks. Four conditions were associated with dependence in four tasks each and posed risks of similar magnitude: knee osteoarthritis, heart disease, congestive heart failure, and chronic obstructive pulmonary disease.

The largest proportions of subjects dependent on others in stair climbing were among elders with hip fracture or congestive heart failure. Increased odds of dependence in stair climbing were also seen for elders with knee osteoarthritis, diabetes, stroke, chronic obstructive pulmonary disease, and depressive symptomatology. The proportion of elders

TABLE 2—Prevalence of Conditions

Condition	No. Affected/ Total n	%	% with at Least One Other Condition
Knee osteoarthritis	318/1726	18.4	50.0
Hip fracture	62/1769	3.5	69.4
Diabetes	110/1758	6.3	69.1
Stroke	122/1769	6.9	71.3
Heart disease	324/1769	18.3	66.4
Congestive heart failure	56/1769	3.2	92.9
Claudication	124/1769	7.0	77.4
Chronic obstructive pulmonary disease	113/1724	6.6	58.4
Depressive symptomatology	152/1668	9.1	57.2
Cognitive impairment	139/1704	8.2	67.6

Note. "Total n's" may not equal 1769 because of missing data.

TABLE 3—Prevalence of Dependence on Human Assistance

Activity	No. Affected/ Total n	%
Stair climbing	101/1733	5.8
Walking a mile	301/1690	17.8
Heavy home chores	531/1712	31.0
Housekeeping	131/1567	8.4
Cooking	69/1558	4.4
Grocery shopping	113/1558	7.3
Carrying bundles	89/1643	5.4

Note. "Total n's" may not equal 1769 because of missing data.

dependent in walking a mile was substantial for every condition. Elders with hip fracture and stroke had the greatest odds of limitation in this activity.

For each of the 10 conditions, more elders were limited in performing heavy home chores than in any other activity. The odds of dependence in performing heavy home chores were particularly striking for elders with stroke, heart disease, and congestive heart failure. Dependence in light housekeeping was not common. The strongest associations were found for elders with hip fracture, stroke, and congestive heart failure. Elders with knee osteoarthritis, heart disease, chronic obstructive pulmonary disease, and depressive symptomatology also had substantial odds of dependence in this activity.

Few elders with any medical condition were functionally limited in cooking. The strongest associations between a medical condition and a functional limitation in cooking were found for elders with stroke or diabetes. The largest proportion of elders who were dependent in grocery shopping were subjects who had sustained a hip fracture at some time prior to examination 18. Dependence in grocery shopping was most strongly associated with stroke. Elders with hip fracture, heart disease, and depressive symptomatology also had increased odds of dependence in grocery shopping. The risk of dependence in carrying bundles was greatest for elders with congestive heart failure, who also constituted the largest proportion of elders limited in this activity. Knee osteoarthritis, stroke, and depressive symptomatology were also strongly associated with dependence in carrying bundles.

The percentage of functional limitation in each task attributable to each medical condition after adjustments were made for age, sex, and comorbidity is found in Table 5. Most conditions account for at least 7% of the disability in at least one activity. Comparisons of the magnitude of these fractions across all activities, however, draw attention to one diagnosis in particular: stroke. Although attributable fractions of similar magnitude in every functional task were found for some other conditions, depressive symptomatology accounts for a substantial proportion of the limitations in each of the five functional tasks. Knee osteoarthritis and heart disease each contribute to limitations in at least four tasks, only two of which (walking a mile and light housekeeping) are the same for both conditions. The proportion of limitations in either of these activities that is associated with knee osteoarthritis is substantially greater than the proportion attributable to heart disease.

Discussion

We identified 10 prevalent medical conditions associated with disability in elders: knee osteoarthritis, hip fracture, diabetes, stroke, heart disease, congestive heart failure, intermittent claudication, chronic obstructive pulmonary disease, depressive symptomatology, and cognitive impairment. Our findings indicate that these 10 medical conditions impose functional limitations on the elderly that are specific to each disease and to each activity. After adjustment for coexisting medical conditions, stroke, depressive symptomatology, hip fracture, knee osteoarthritis, heart disease, congestive heart failure, and chronic obstructive pulmonary disease demonstrated significant and substantial associations with limitations in at least four of the seven functional activities studied. The largest proportions of disability were attributable to knee osteoarthritis, heart disease, depressive symptomatology, and stroke.

The results of this study corroborate the findings of other studies. Despite its low prevalence, stroke, which limits a broad array of functional activities, is one of the main disabling conditions of the elderly. Our sample included only noninstitutionalized survivors of stroke and excluded those who may have been functionally incapacitated by stroke to the point of needing nursing home placement. Thus, with respect to all elders, the disabling effects of stroke may be even greater than our results indicate. Previous studies, including those done at Framingham,[19,20] have indicated a relationship between stroke and disability in the total number of activities that were limited. The data presented above clarify the magnitude of this relationship. Our study also demonstrates the global effect of stroke on function, which follows from the pattern of upper- and lower-extremity neuromuscular impairments typically seen in the patient with stroke. Similarly, our findings support the observations of Harris et al.[15] and Guralnik and Kaplan[8] regarding arthritis and cardiovascular disease and their relationship to disability in the elderly.

The conclusions on the relationship of cardiovascular diseases to physical disability in the elderly that are supported by our data are somewhat differ-

TABLE 4—Odds of Disability, Adjusted for Age, Sex, and Comorbidity, by Condition and by Activity

	Stair Climbing		Walking a Mile		Heavy Home Chores		Housekeeping		Cooking		Grocery Shopping		Carrying Bundles	
	Proportion Disabled (%)	Adjusted OR (95% CI)	Proportion Disabled (%)	Adjusted OR (95% CI)	Proportion Disabled (%)	Adjusted OR (95% CI)	Proportion Disabled (%)	Adjusted OR (95% CI)	Proportion Disabled (%)	Adjusted OR (95% CI)	Proportion Disabled (%)	Adjusted OR (95% CI)	Proportion Disabled (%)	Adjusted OR (95% CI)
Knee osteoarthritis	32/315 (10.2)	1.98 (1.14, 3.43)	95/305 (31.2)	1.91 (1.38, 2.63)	106/311 (34.4)	0.98 (0.74, 1.30)	40/286 (14.0)	2.09 (1.29, 3.39)	15/285 (5.3)	1.02 (0.48, 2.19)	27/277 (9.8)	1.05 (0.58, 1.89)	28/296 (9.5)	2.02 (1.15, 3.54)
Hip fracture	13/58 (22.4)	3.39 (1.56, 7.34)	28/56 (50.0)	2.12 (1.22, 3.67)	37/58 (63.8)	1.65 (1.02, 2.66)	17/57 (29.8)	2.49 (1.21, 5.12)	9/58 (15.5)	2.19 (0.70, 6.90)	15/56 (26.8)	2.30 (1.06, 4.99)	13/58 (22.4)	2.01 (0.85, 4.73)
Diabetes	15/109 (13.8)	2.90 (1.45, 5.77)	31/104 (29.8)	1.70 (1.05, 2.73)	46/105 (43.8)	1.33 (0.89, 1.98)	11/100 (11.0)	1.43 (0.68, 3.04)	8/103 (7.8)	2.33 (1.01, 5.35)	10/101 (9.9)	1.62 (0.76, 3.45)	9/102 (8.8)	1.82 (0.79, 4.21)
Stroke	20/118 (17.0)	2.82 (1.40, 5.70)	50/116 (43.1)	2.27 (1.46, 3.53)	74/115 (64.4)	1.93 (1.35, 2.75)	25/109 (22.9)	2.95 (1.59, 5.45)	17/108 (15.7)	2.48 (1.12, 5.49)	26/111 (23.4)	3.63 (1.98, 6.67)	17/108 (15.7)	2.93 (1.39, 6.20)
Heart disease	28/315 (8.9)	1.23 (0.67, 2.24)	85/308 (27.6)	1.49 (1.06, 2.09)	167/316 (52.9)	1.95 (1.51, 2.51)	39/289 (13.5)	1.82 (1.10, 3.01)	21/291 (7.2)	1.41 (0.71, 2.80)	33/294 (11.2)	1.80 (1.06, 3.06)	22/296 (7.4)	1.10 (0.58, 2.11)
Congestive heart failure	12/56 (21.4)	2.76 (1.19, 6.38)	21/52 (40.4)	1.52 (0.81, 2.84)	40/55 (72.7)	2.06 (1.27, 3.35)	14/50 (28.0)	2.52 (1.14, 5.58)	6/49 (12.2)	1.61 (0.46, 5.72)	7/51 (13.7)	1.36 (0.49, 3.79)	9/51 (17.7)	2.89 (1.17, 7.14)
Claudication	11/121 (9.1)	1.39 (0.63, 3.08)	42/120 (35.0)	1.96 (1.25, 3.09)	52/119 (43.7)	1.14 (0.78, 1.68)	12/114 (10.5)	1.00 (0.47, 2.11)	7/115 (6.1)	1.43 (0.60, 3.43)	11/114 (9.7)	0.93 (0.42, 2.07)	9/116 (7.8)	1.43 (0.61, 3.34)
Chronic obstructive pulmonary disease	12/111 (10.8)	2.34 (1.08, 5.07)	32/111 (28.8)	1.86 (1.15, 3.00)	52/109 (47.7)	1.59 (1.09, 2.34)	14/106 (13.2)	2.31 (1.16, 4.62)	6/106 (5.7)	0.88 (0.29, 2.61)	11/106 (10.4)	1.38 (0.61, 3.10)	56/135 (41.5)	2.14 (0.91, 5.05)
Depressive symptomatology	22/149 (14.8)	2.35 (1.27, 4.33)	58/146 (39.7)	1.98 (1.34, 2.90)	74/147 (50.3)	1.36 (0.97, 1.88)	31/136 (22.8)	2.38 (1.39, 4.09)	8/135 (5.9)	1.44 (0.57, 3.69)	22/130 (16.9)	2.31 (1.30, 4.11)	17/128 (13.3)	2.53 (1.38, 4.64)
Cognitive impairment	15/136 (11.0)	0.82 (0.33, 2.06)	39/136 (28.7)	0.99 (0.60, 1.61)	56/135 (41.5)	0.96 (0.64, 1.44)	23/123 (18.7)	1.21 (0.59, 2.47)	16/131 (12.2)	1.37 (0.59, 3.20)	22/130 (16.9)	1.54 (0.79, 3.02)	17/128 (13.3)	1.65 (0.76, 3.56)

Note. OR = odds ratio; CI = confidence interval.

TABLE 5—Percentage of Disability Attributable to Specific Conditions after Adjustment for Age, Sex, and Comorbidity, by Condition and Activity

Activity	Knee Osteo-arthritis	Hip Fracture	Diabetes	Stroke	Heart Disease	Congestive Heart Failure	Claudication	Chronic Obstructive Pulmonary Disease	Depressive Symptomatology	Cognitive Impairment
Stair climbing	16.7	9.1	9.8	12.8	5.2	7.6	3.1	7.4	15.4	. . .[a]
Walking a mile	15.4	4.9	4.3	9.3	9.3	2.4	6.6	5.2	10.4	. . .[a]
Heavy home chores	. . .[a]	2.8	2.2	6.7	15.3	3.9	1.2	3.8	4.0	. . .[a]
Housekeeping	16.7	7.8	2.6	12.6	13.4	6.5	0.0	6.5	15.9	3.4
Cooking	0.5	7.1	6.8	14.7	8.9	3.3	3.1	. . .[a]	4.3	7.3
Grocery shopping	1.2	7.5	3.5	16.7	13.0	1.6	. . .[a]	4.5	15.4	7.8
Carrying bundles	16.6	7.3	4.6	12.6	2.3	6.6	3.0	4.0	16.6	8.6

[a]Adjusted odds ratio < 1.

ent from those reached by other investigators at Framingham.[14] We found substantial effects for cardiovascular disease including congestive heart failure. Previous analyses of physical disability in the Framingham cohort may have been unable to detect relationships when the cohort was younger and therefore less disabled. Furthermore, our analyses indicate that there is some task-specificity in the disability associated with cardiovascular disease. Summating the number of functionally limited activities may obscure the relationship between some cardiovascular diseases and limitations in particular activities.

The diminution of the effects of cognitive impairment after adjustment for comorbidity is striking. Although the relationship between cognitive and functional decline has been well accepted, it is clear that much of the disability associated with cognitive impairment among noninstitutionalized individuals may be due to other causes. It is also possible that the functional measures we chose do not capture the specific tasks that are most affected by cognitive impairment in a community-dwelling population. In contrast, however, the effects of depressive symptomatology on function remain even after adjustment for the comorbidities included in our study. Previously, Wells and colleagues found that adults with a current depressive disorder or depressive symptoms had worse physical, social, and role functioning than their peers without such conditions.[35] Our finding highlights the contributions of psychological factors to physical functional limitations in the elderly. Physical disability is not merely the end result of physical conditions. Successful performance of many functional activities, such as housekeeping, requires motivation and cognitive abilities as well as physical ones.

It is also important to comment on two of the conditions for which we were unable to find a relationship between disease and disability, namely cancer and visual impairment. It was surprising to us that prevalent cancers were not related to functional limitations in the elderly. Satariano et al., for example, have found difficulty in activities requiring primarily upper-extremity strength among women with breast cancer aged 55 through 74, but not among those aged 75 through 84.[21] Although our five complex activities of daily living required the use of both upper and lower extremities, we could not find any consistent relationship between cancer and disability in our preliminary analyses. The mean age of our sample approximates the age of the group in Satariano's sample who were no more disabled than the control subjects. It is quite possible that those elders in Framingham with the most disabling forms of cancer had died of it or did not participate in the study; therefore an older cohort such as ours may include only elders with less disabling cancer. Although visual impairment may increase the risk of falls and other accidents among the elderly, it does not appear to be related to functional limitations in the tasks we evaluated. It is also possible that the criterion of visual impairment we used (best-corrected vision of 20/40) may not be severe enough to identify those elders whose functional limitations may be associated with impaired ability to see.

Verbrugge and her colleagues reported on the relative impact of comorbid conditions on functional limitations in all adults aged 55 years or older included in the Supplement on Aging to the 1984 National Health Interview Survey.[36] Considering each disease separately and adjusting for age, sex, and race, Verbrugge et al. found that highly prevalent conditions generally had little impact on an elder's difficulty in performing functional activities. Conversely, several low-prevalence conditions were strongly associated with increases in the total number of gross mobility tasks and basic and instrumental activities of daily living that elders performed with difficulty or were unable to do. Arthritis was shown to be the sole exception to this general conclusion, as it was a highly prevalent disease with a moderate correlation with difficulty in performing an overall number of physical functional tasks. Our analyses corroborate and extend the conclusions of Verbrugge et al. However, it should be noted that our definition of medically verified arthritis was limited to the knee. If we had included hand, wrist, and hip arthritis, the functional impact of "all arthritis," the definition used by Verbrugge et al. for self-reported arthritis and in calculating the proportion of disability attributable to arthritis, would have been even greater.

The conclusions of our study are limited in several ways. Our data were collected at the Framingham study site. Therefore, some individuals may have been excluded because they were functionally limited by their medical conditions to the point of being unable to come to the site. The nonparticipation of these individuals would weaken the association of disease and disability in our results.

Another limitation of this study may be the definitions we chose to use for each of the medical conditions, despite the accuracy of data verified medically. As we have shown for knee osteoarthritis, the risks of disability change with the stringency of the definition of the disease.[2] We can assume that the risks of disability for any other condition we studied would change if the criteria for the condition were modified. When confronted with the possibility of defining a condition in several ways, we chose conservative criteria that we believed would also be clinically meaningful. Furthermore, our study was concerned only with the association between a single condition and physical disability after all other medical conditions were controlled. We did not examine how various combinations of medical conditions might be associated with increased risk for disability in the elderly. Our understanding of disability in the elderly should not, however, be confined merely to the study of disease and disability. Future research must also account for the contributions of psychological and social factors—for example, personal motivation and social support—to specific physical functional limitations.

Our study is also limited by the functional tasks that were available for study. Although these items are traditionally used in broad epidemiological studies of disability in the elderly, they are only a very small sample of functional activities. Other data, not available at Framingham, might be able to demonstrate similarly strong associations between our 10 target conditions and other functional tasks. A change in dependent variables might also yield substantial odds of disability for some of the conditions that did not appear strongly associated with functional limitations in our study. For example, it is very likely that the odds of functional limitation in elders with cognitive impairment would have been greater if we had chosen basic hygiene and self-care activities of daily living as measures of disability.

The results of this study have public policy implications. The emphasis in geriatric health care and prevention has only recently been on functional morbidity. The relationship between stroke and functional limitations in the elderly is broad, involving a host of activities. Reductions in stroke and its aftereffects should yield improvements in the functional abilities of elders in a variety of tasks. Similar gains in physical function might be realized through prevention of depressive symptomatology and hip fracture. In contrast, the relationships of knee osteoarthritis and heart disease to function are more closely tied to specific activities. Thus, practitioners must carefully link a functional deficit to its most likely medical cause to diminish the effects of these diseases on function. Clinically, as demonstrated by the high degree of medical comorbidity in our subjects, these medical conditions are likely to be coexistent. Successful evaluation and treatment of an elderly patient will require a careful examination of the independent contributions of each of the patient's medical diagnoses to his or her physical functional status.

In summary, we investigated the effects of 10 medical conditions on seven functional limitations in the elderly. Only one condition, stroke, contributed to limitations in all seven activities. The proportions of disability in the elderly attributable to knee osteoarthritis, stroke, heart disease, and depressive symptomatology were greater than the proportions of disability attributable to other conditions. Efforts to reduce the effects of certain medical conditions on the functional status of the elderly will be enhanced by our understanding of the disease- and task-specificity of these relationships. □

Acknowledgments

This work was supported by NIH-NIA grants KO1AG00567 (Dr Guccione) and RO1AG07861 (Dr Felson) and NIH-NIAMS Multipurpose Arthritis Center grant P60AR20613.

We are grateful to Roy C. Milton, PhD, Biometry and Epidemiology Program, National Eye Institute, National Institutes of Health, for his assistance with this study.

References

1. Guralnik JM, LaCroix AZ, Everett DF, Kovar MG. Aging in the eighties: the prevalence of comorbidity and its association with disability. *Adv Data Vital Health Stat.* May 26, 1989; no. 170. DHHS publication no. PHS 89-1250.
2. Guccione AA, Felson DT, Anderson JJ. Defining arthritis and measuring functional status in elders: methodological issues in the study of disease and physical disability. *Am J Public Health.* 1990;80: 945–949.
3. Yelin EH, Spitz PP. Transitions in health status among community-dwelling elderly people with arthritis. *Arthritis Rheum.* 1990;33:1205–1215.
4. Verbrugge LM, Gates DM, Ike RW. Risk factors for disability among US adults with arthritis. *J Clin Epidemiol.* 1991;44: 167–182.
5. Verbrugge LM, Lepkowski JM, Konkol LL. Levels of disability among US adults with arthritis. *J Gerontol.* 1991;46:S71–S83.
6. Kaplan GA. Epidemiologic observations on the compression of morbidity. *J Aging Health.* 1991;3:155–171.
7. Mor V, Murphy J, Masterson-Allen S, et al. Risk of functional decline among well elders. *J Clin Epidemiol.* 1989;42:895–904.
8. Guralnik JM, Kaplan GA. Predictors of healthy aging: prospective evidence from the Alameda County study. *Am J Public Health.* 1989;79:703–708.
9. Magaziner J, Simonsick EM, Kashner TM, Hebel JR, Kenzora JE. Predictors of functional recovery one year following hospital discharge for hip fracture: a prospective study. *J Gerontol.* 1990;45: M101–107.
10. Lavsky-Shulan M, Wallace RB, Kohout FJ, Lemke JH, Morris MC, Smith IM. Prevalence and functional correlates of low back pain in the elderly: the Iowa 65+ Rural Health Study. *J Am Geriatr Soc.* 1985;33:23–28.
11. Barrett-Connor E, Wingard DL. Heart disease risk factors as determinants of dependency and death in an older cohort. *J Aging Health.* 1991;3:247–261.
12. Pinsky JL, Branch LG, Jette AM, et al. The Framingham Disability Study: relationship of disability to cardiovascular risk factors among persons free of diagnosed cardiovascular disease. *Am J Epidemiol.* 1985;122:644– 656.
13. Nickel JT, Chirikos TN. Functional disability of elderly patients with long-term coronary heart disease: a sex-stratified analysis. *J Gerontol.* 1990;45:S60–S68.
14. Pinsky JL, Jette AM, Branch LG, Kannel WB, Feinleib M. The Framingham Disability Study: relationship of various coronary heart disease manifestations to disability in older persons living in the community. *Am J Public Health.* 1990;80: 1363–1368.
15. Harris T, Kovar MG, Suzman R, Kleinman JC, Feldman JJ. Longitudinal study of physical ability in the oldest-old. *Am J Public Health.* 1989;79:698–702.
16. Cook NR, Evans DA, Scherr FE, et al. Peak expiratory flow rate in an elderly population. *Am J Epidemiol.* 1989;130:66–78.
17. Mossey JM, Knott K, Craik R. The effects of persistent depressive symptoms on hip fracture recovery. *J Gerontol.* 1990;45: M163–M168.
18. Berkman LF, Berkman CS, Kasl S, et al. Depressive symptoms in relation to physical health and functioning in the elderly. *Am J Epidemiol.* 1986;124:372–388.
19. Kelly-Hayes M, Wolf PA, Kannel WB, Sytkowski P, D'Agostino RB, Gresham GE. Factors influencing survival and need for institutionalization following stroke: the Framingham Study. *Arch Phys Med Rehabil.* 1988;69:415–418.
20. Jette AM, Pinsky JL, Branch LG, Wolf PA, Feinleib M. The Framingham Disability Study: Physical disability among community-dwelling survivors of stroke. *J Clin Epidemiol.* 1988;41:719–726.
21. Satariano WA, Ragheb NE, Branch LG, Swanson GM. Difficulties in physical

functioning reported by middle-aged and elderly women with breast cancer: a case control comparison. *J Gerontol.* 1990;45: M3–M11.
22. Scherr PA, Albert MS, Funkenstein HH, et al. Correlates of cognitive function in an elderly community population. *Am J Epidemiol.* 1988;128:1084–1101.
23. Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic conditions: results from the Medical Outcomes Study. *JAMA.* 1989;262:907–913.
24. Lair T, Lefkowitz D. *Mental Health and Functional Status of Residents of Nursing and Personal Care Homes.* Rockville, Md: Public Health Service; September 1990. DHHS publication PHS 90-3470. National Medical Expenditure Survey Research Findings 7, Agency for Health Care Policy and Research.
25. Kellgren JH, Lawrence JS. *Atlas of Standard Radiographs: The Epidemiology of Chronic Rheumatism.* Vol 2. Oxford, England: Blackwell Scientific Publications; 1963.
26. Samet JM. Definitions and methodology in COPD research. In: Hensley MJ, Saunders NA, eds. *Clinical Epidemiology of Chronic Obstructive Pulmonary Disease.* New York, NY: Marcel Dekker Inc; 1989:1–22.
27. Bachman DL, Wolf PA, Linn R, et al. Prevalence of dementia and probable senile dementia of the Alzheimer type in the Framingham Study. *Neurology.* 1992; 42:115–119.
28. Folstein MF, Folstein SE. Syndromes of altered mental state. In: Hazzard WR, Andres R, Bierman EL, Blass JP, eds. *Principles of Geriatric Medicine and Gerontology.* 2nd ed. New York, NY: McGraw-Hill Inc; 1990:1089–1101.
29. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *J Appl Psychol Meas.* 1977;1:385–401.
30. Rosow I, Breslau N. A Guttman health scale for the aged. *J Gerontol.* 1966;21:556–559.
31. Nagi SZ. An epidemiology of disability among adults in the United States. *Milbank Mem Fund Q.* 1976;54:439–467.
32. Branch LG, Jette AM. The Framingham Disability Study, I: social disability among the aging. *Am J Public Health.* 1981;71: 1202–1210.
33. Kelly-Hayes M, Jette AM, Wolf PA, et al. Functional limitations and disability among elders in the Framingham Study. *Am J Public Health.* 1992;82:841–845.
34. Greenland S. Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. *Stat Med.* 1984;3:131–141.
35. Wells KB, Stewart A, Hays RD, et al. The functioning and well-being of depressed patients: results from the Medical Outcomes Study. *JAMA.* 1989;262:914–919.
36. Verbrugge LM, Lepkowski JM, Imanaka Y. Comorbidity and its impact on disability. *Milbank Q.* 1989;67:450–484.

APPENDIX A—Medical Conditions Included in Preliminary Analyses

- Musculoskeletal
  - Knee osteoarthritis
  - Hip fracture
- Endocrine
  - Obesity
  - Diabetes
- Cardiovascular
  - Hypertension
  - Stroke excluding transient ischemic attack
  - Angina pectoris
  - Myocardial infarction
  - Coronary insufficiency
  - Heart disease
  - Congestive heart failure
  - Intermittent claudication
- Pulmonary
  - Dyspnea on exertion
  - Chronic obstructive pulmonary disease
  - Chronic obstructive pulmonary disease with dyspnea on exertion
- Oncologic
  - Cancers except nonmelanoma skin cancers
- Psychological
  - Depressive symptomatology
  - Cognitive impairment
- Visual impairment
  - Cataract
  - Visual acuity

APPENDIX B—Definitions of Medical Conditions

Knee osteoarthritis: Symptomatic osteoarthritis defined as pain in or around the knee lasting at least a month and at least grade 2 x-ray changes or asymptomatic osteoarthritis at least grade 3 on x-ray.

Hip fracture: Any hip fracture sustained prior to exam 18 visit.

Diabetes: Currently treated diabetes based on use of oral hypoglycemics or insulin.

Stroke: Sudden onset of a focal neurological deficit lasting more than 24 hours prior to exam 18 visit, excluding transient ischemic attacks.

Heart disease: Either angina pectoris or definite myocardial infarction on electrocardiogram any time prior to exam 18 visit.

Congestive heart failure: Condition meeting a minimum of two major criteria (e.g., paroxysmal nocturnal dyspnea or orthopnea, neck vein distension, cardiomegaly, acute pulmonary edema, rales, or S_3 gallop) or one major criterion and two minor criteria (e.g., bilateral ankle edema, dyspnea on exertion, tachycardia, or night cough).

Claudication: Intermittent claudication diagnosed by Framingham physicians.

Chronic obstructive pulmonary disease: $FEV_1 \div FVC < 60\%$ by spirometry at exam 16 or 17.[a]

Depressive symptomatology: A score of more than 4 of 10 on a modified CES-D scale with 10 questions.

Cognitive impairment: A score of less than 24 on the Folstein Mini-Mental Status Examination.

[a] FEV_1 = forced expiratory volume in 1 second; FVC = forced vital capacity.

US Dietary Patterns Associated with Fat Intake: The 1987 National Health Interview Survey

ABSTRACT

Objectives. This research used food frequency data to investigate dietary patterns associated with fat intake.

Methods. Data from the 1987 National Health Interview Survey of 20 143 adults were used to determine correlations between fat (adjusted for kilocalories) and both nutrient and food group intakes. Median food and nutrient intakes were determined within quartiles of percentage of kilocalories from fat.

Results. Intakes of vegetables, fruits, cereals, fish/chicken, low-fat milk, alcoholic beverages, vitamin C, percentage of kilocalories from carbohydrates, carotenoids, folate, dietary fiber, carbohydrates, and vitamin A decreased as percentage of kilocalories from fat increased. Intakes of salty snacks, peanuts, processed and red meats, whole milk and cheese, desserts, eggs, fried potatoes, table fats, cholesterol, vitamin E, sodium, protein, and energy increased with percentage of kilocalories from fat. Results by demographic subgroups showed few differences from those found in the total population.

Conclusions. Fat intake is consistently associated with specific dietary patterns. Such patterns need to be evaluated concurrently in studies of diet and chronic disease. (*Am J Public Health.* 1994;84:359–366)

Amy F. Subar, PhD, MPH, RD, Regina G. Ziegler, PhD, MPH, Blossom H. Patterson, MA, Giske Ursin, MD, and Barry Graubard, PhD

Introduction

In the last 2 decades, there has been increasing interest in the role of dietary fat in chronic disease. The consensus among major scientific organizations is that fat intake should be reduced below current levels in the United States to reduce the risk of heart disease, cancer, and obesity[1–6]; the most common recommendation is that individuals reduce their fat intake to 30% of calories (kilocalories) or less.[3,4,6] Many of these organizations recommend concurrent dietary guidelines that include increased fruit and vegetable intake to five daily servings,[2,3,6] increased intake of whole grains and dietary fiber,[1,3–6] and reduced sodium intake.[1–6] When individuals change their fat intake, however, a variety of other dietary changes may occur to balance caloric intake. Food selections are based on food preference and availability, budget, social environment, and other factors. A low-fat diet can be implemented in a variety of ways: increasing intake of fruits, vegetables, complex carbohydrates, and/or low-fat dairy products and decreasing intake of meats, table fats, whole milk and cheese, and/or baked goods. These choices may produce quite different dietary patterns, many of which could constitute a diet low in fat. Two recent studies[7,8] examined dietary patterns and fat intake in small numbers of women (172 and 17, respectively) volunteering for clinical trials designed to reduce fat intake. The purpose of the present study is to quantify the relationships between dietary patterns and diets either low or high in fat among 20 143 adults interviewed in the nationally representative 1987 National Health Interview Survey (NHIS).

Methods

The 1987 NHIS, conducted by the National Center for Health Statistics, was administered throughout calendar year 1987 to 44 123 individuals and had a response rate of 82%. The survey was designed to represent US households in the 48 contiguous states.[9] One adult per household was randomly selected to be administered the Cancer Risk Factor Supplement, developed and funded by the National Cancer Institute. Details regarding the methodology of the Cancer Risk Factor Supplement and the collection of dietary data can be found elsewhere.[10] Briefly, in a split-sample design, either the cancer control or the epidemiology questionnaire was administered. Door-to-door, in-home interviews of non-institutionalized individuals were conducted by experienced Bureau of the Census interviewers. For respondents not found at home after repeated visits, interviews were obtained by telephone whenever possible (approximately 16%). Hispanics were oversampled, and a Spanish translation of the interview was used when necessary.

Amy F. Subar, Blossom H. Patterson, and Barry Graubard are with the Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Md. Regina G. Ziegler is with the Division of Cancer Etiology, National Cancer Institute. Giske Ursin is with the Department of Preventive Medicine, School of Medicine, University of Southern California, Los Angeles.

Requests for reprints should be sent to Amy F. Subar, PhD, MPH, RD, National Cancer Institute, Division of Cancer Prevention and Control, Surveillance Program, Applied Research Branch, 9000 Rockville Pike, EPN 313, Bethesda, MD 20892.

This paper was accepted June 3, 1993.

TABLE 1—Distribution of Demographic Groups in Quartiles of Percentage of Kilocalories from Fat: National Health Interview Survey, 1987

Demographic Group	Sample Size	Quartile %[a]			
		1	2	3	4
Sex					
Male	8385	25.7	25.1	25.0	24.2
Female	11 758	24.1	24.5	25.0	26.4
Age, y					
18–34	7462	22.7	25.8	26.8	24.6
35–49	5269	23.8	23.7	25.8	26.7
50–64	3604	27.2	24.5	22.5	25.8
65+	3808	29.3	24.5	22.0	24.2
Race/ethnicity					
White	16 065	23.6	24.9	25.2	26.4
Black	2748	25.0	22.5	26.5	25.9
Hispanic	1330	39.7	26.7	20.7	12.9
Education, y					
<12	4725	25.2	23.0	24.3	27.5
12	7632	21.4	24.0	26.5	28.2
>12	7731	28.3	26.6	23.9	21.2
Unknown	55	27.8	31.9	21.1	19.2
Income, dollars					
<14 000	5476	25.8	22.9	25.5	25.8
14 000–29 999	6150	23.2	23.9	25.5	27.4
30 000–49 999	4264	22.8	26.1	26.0	25.1
≥50 000	2112	30.4	27.7	21.4	20.5
Unknown	2141	25.4	24.3	24.9	25.4

[a]Quartile cutpoints for percentage of kilocalories from fat were 33.9%, 38.8%, and 43.3%.

The dietary interview, consisting of a 59-item food frequency questionnaire developed by Block et al.,[11] was administered to 22 080 adults and lasted an average of 17 minutes. Individuals were asked to report the number of times per day, week, month, or year they consumed each food item. Portion size was asked for each food item (small, medium, or large), and age- and sex-specific portion sizes were used to estimate nutrient intake.[12] The 59 food items included on the questionnaire were selected to represent the foods that were the major contributors of nutrients in adult diets based on analyses of nutrient sources[13,14] from 24-hour dietary recall data collected in the second National Health and Nutrition Examination Survey (NHANES II) during 1976 through 1980. The longer 98-item questionnaire on which this instrument was based was validated in several studies[15–18] and has shown good ranking of individuals and reasonable estimates of mean nutrient intake. Correlations between nutrients estimated by multiple days of diet records or the 59-item questionnaire were comparable to those found with the longer version.[11] Of the 22 080 adults who were administered the questionnaire, approximately 6% were excluded from analysis because of coding, interviewer, or respondent error, as were individuals reporting a racial or ethnic identification other than White, Black, or Hispanic. Data on 20 143 individuals were analyzed.

Nutrient estimates were calculated with the database and software developed by Block et al. based on the NHANES II food database.[12] In comparison with the 98-item questionnaire, the 59-item instrument tends to underestimate absolute intake of macronutrients and some micronutrients but may overestimate vitamins A and C. However, percentage of kilocalories from macronutrients is considered accurate.[19] Because of this concern about estimates of absolute intake, relative intake and dietary patterns are emphasized. Adjustment coefficients[19] were not used. The 59 food items were categorized into groups (not always mutually exclusive) based on nutritional, functional, and botanical similarities, and total servings per week were calculated (a description of the food groups is available on request).

Individuals were categorized by their percentage of kilocalories from fat, a measure widely recognized as of interest to public health. First, distributions were examined separately by sex, by race/ethnicity, by age (19 through 34, 35 through 49, 50 through 64, 65+ years), and by eight sex-age groups. Distributions for all subgroups were nearly identical, except for Hispanics, who had somewhat lower percentages of intake of kilocalories from fat. Thus, subjects were assigned to quartiles of percentage of kilocalories from fat based on the total population. Medians for nutrient and food group intake were examined by quartile for all age–sex groups and by race/ethnicity, education (<12, 12, >12 years), and household income (<$14 000, $14 000 through $29 999, $30 000 through $49 999, ≥$50 000). Three methods were used in examining correlations between fat intake and nutrient/food group intake controlling for total caloric intake. The first method involves Pearson product moment correlations between absolute nutrient intake and percentage of kilocalories from fat, a commonly used procedure yielding results that can be compared with those of other studies. The second method computes Pearson correlations between age- and sex-specific percentiles of nutrient intake and percentage of kilocalories from fat. The use of percentiles adjusts nutrient intake for age and sex differences. In addition, this procedure has the advantage of not being unduly influenced by extreme values of nutrient intake. The third method uses partial correlations between the nutrient/food group and total fat intake adjusting for kilocalories, age, and sex. All correlations and median intake values were calculated by using sample weights in SAS[20] and SUDAAN.[21]

Results

The mean percentage intake of kilocalories from fat for the entire NHIS population was 38.4% (range = 6.0% to 71.6%). Weighted quartile cutpoints for percentage of kilocalories from fat were 33.9%, 38.8%, and 43.3%. Table 1 shows the number and percentage of individuals in each quartile by demographic subgroup. Hispanics were more likely to be in the lowest quartile than were Whites or Blacks. The percentage of individuals in the lowest quartile increased consistently with age, while the percentage in the third quartile de-

creased from the youngest age groups to those 50 or older. Analyses by education and income showed that both the wealthiest and the most educated were more likely to consume a low percentage of kilocalories from fat.

Table 2 shows that the correlation coefficients between nutrient intake and fat adjusted for kilocalories obtained with the three methods described above were quite similar. Exceptions were oleic acid, saturated fat, and carbohydrate, for which the partial correlations with fat adjusted for calories were noticeably larger than either correlation with percentage of kilocalories from fat. On the basis of the overall similarity of the three approaches, additional analyses used simple correlations.

The correlations for macronutrients (Table 2) show that total fat and fat constituents (oleic acid, saturated fat, linoleic acid, and cholesterol) were most highly positively correlated with percentage of kilocalories from fat (r's ranging from −.29 to .42). Protein (r = .13) and carbohydrate (r = −.19) had low correlations, with protein increasing and carbohydrate decreasing with increased percentage of kilocalories from fat. The percentages of kilocalories from fat and from carbohydrate were strongly negatively correlated (r = −.73), compared with only a slight positive correlation for percentage of kilocalories from protein (r = .08). Energy intake had a low positive correlation (r = .07) with percentage of kilocalories from fat.

Of the micronutrients, vitamin C had the largest negative correlation with percentage of kilocalories from fat (r = −.32). Folate, dietary fiber, and carotene also had negative correlations (r's of approximately −0.2), and vitamin E and sodium had positive correlations of similar magnitude. Correlations were less than ±.10 for all other nutrients. Simple correlations with percentage of kilocalories from fat for all nutrients in Table 2 were also determined by sex, four age groups, three race groups, three education groups, and four income groups to investigate differences from the total population correlations. Of the 400 possible correlations, only 7 (2%) varied by ±.10 from the population correlation, and no consistent patterns emerged.

Table 3 shows the absolute median intake of nutrients by percentage of kilocalories from fat quartile and the ratio of medians in quartile 4 to medians in quartile 1. Nutrients are listed in descending order by ratio. The similar order of the nutrients in Table 2 (by correlation coefficient) and Table 3 (by ratio) suggests relative consistency in their associations with percentage of kilocalories from fat. Energy intake was 14% greater for individuals in the highest compared with the lowest quartile. For fat and its constituents, the differences were more dramatic, with intakes of oleic acid, saturated fat, linoleic acid, and total fat almost twice as great, and cholesterol two thirds as great in the fourth as in the first quartile. Similar to energy, intake of protein was 17% greater in the fourth than the first quartile. Intake of total carbohydrate in the fourth quartile was about one quarter of that in the first, while percentage intake of kilocalories from carbohydrate was three quarters of that in the first.

Results for micronutrients and fiber showed a trend for intakes of vitamin C, dietary fiber, folate, and carotene, all of which are found primarily in fruits and vegetables, to be lowest in the highest quartile. The gradient was most striking for vitamin C, with intake in the fourth quartile about 60% of that in the first. Intakes of dietary fiber, folate, and carotene among those in the highest quartile were all about three fourths of those in the lowest. In contrast, intakes of sodium and vitamin E were higher in the fourth quartile than in the first (1.24 and 1.33 times as great, respectively). Other nutrients varied little by quartile. While absolute intake of nutrients varied by age and sex (data not shown), the overall pattern of nutrient intake by percentage of kilocalories from fat differed little when ratios were compared (the largest ratio range [1.17 to 2.03] involved saturated fat). Examination of comparable ratios by education, income, and race showed that of the 250 ratios calculated, 2 differed by more than ±.20 from the population ratio.

Table 4 provides correlations between intake of food groups (number of servings per week) and percentage of kilocalories from fat. Results for all three methods were essentially the same, with the exception of alcoholic bever-

TABLE 2—Correlations between Fat and Nutrient Intake: National Health Interview Survey, 1987

Nutrient	Simple Correlation[a]	Percentile Correlation[b]	Partial Correlation[c]
Fat	.42	.55	1.00
Oleic acid	.42	.55	.95
Saturated fat	.40	.51	.87
Linoleic acid	.37	.47	.55
Cholesterol	.29	.36	.36
Vitamin E	.20	.26	.27
Sodium	.17	.21	.30
Protein	.13	.16	.24
Protein, % kcal	.08	.08	.14
Energy	.07	.12	.14
Iron	.02	.03	.03
Phosphorus	−.001	−.01	−.10
Retinol	−.004	.01	−.01
Calcium	−.02	−.02	−.04
Niacin	−.02	−.02	−.17
Thiamin	−.04	−.05	−.06
Riboflavin	−.04	−.05	−.11
Potassium	−.06	−.08	−.14
Vitamin A	−.08	−.08	−.07
Carotene	−.15	−.14	−.13
Carbohydrate	−.19	−.23	−.62
Fiber	−.22	−.22	−.24
Folate	−.24	−.26	−.35
Vitamin C	−.32	−.30	−.32
Carbohydrate, % kcal	−.73	−.70	−.60

[a]Pearson correlation of percentage of kilocalories from fat with absolute intake.
[b]Pearson correlation of percentage of kilocalories from fat with age- and sex-specific percentile of intake.
[c]Correlation of grams of fat with absolute intake (controlling for age, sex, and kilocalories).

TABLE 3—Median Nutrient Intake, by Percentage of Kilocalories from Fat Quartile: National Health Interview Survey, 1987

Nutrient	Quartile %[a]				Ratio[b]
	1	2	3	4	
Oleic acid, g	15.9	21.9	26.0	29.7	1.87
Saturated fat, g	16.5	23.0	26.9	30.6	1.85
Linoleic acid, g	8.0	11.0	13.1	14.9	1.86
Fat, g	44.5	61.1	71.4	82.1	1.84
Cholesterol, mg	200	257	302	331	1.66
Fat, % kcal	30.1	36.5	40.9	46.3	1.54
Vitamin E, alpha tocopherol equivalents	4.9	5.7	6.2	6.5	1.33
Sodium, mg	1948	2204	2360	2414	1.24
Protein, g	54.2	61.0	63.7	63.4	1.17
Energy, kcal	1381	1509	1571	1575	1.14
Protein, % kcal	15.6	16.1	16.2	16.2	1.04
Phosphorus, mg	1008	1061	1071	1021	1.01
Iron, mg	9.4	9.9	10.1	9.5	1.01
Thiamin, mg	1.0	1.0	1.0	1.0	1.00
Retinol, mg	595	632	627	583	0.98
Niacin, mg	14.2	14.6	14.9	13.9	0.98
Calcium, mg	680	705	709	655	0.96
Riboflavin, mg	1.5	1.6	1.6	1.4	0.93
Potassium, mg	2231	2234	2233	2067	0.93
Vitamin A, retinol equivalents	1098	1074	1048	950	0.87
Carbohydrate, g	172	168	162	141	0.82
Dietary fiber, g	10.4	9.6	9.0	7.9	0.76
Folate, μg	262	239	224	194	0.74
Carotene, μg	2494	2232	2108	1848	0.74
Carbohydrate, % kcal	51.0	45.4	41.7	36.3	0.71
Vitamin C, mg	134	114	100	80	0.59

[a]Quartile cutpoints were 33.9%, 38.8%, and 43.3%.
[b]Ratio of quartile 4 to quartile 1.

ages, and simple correlations were used in further analyses. Servings of table fats and "all meats" represented the food group most highly positively correlated with percentage of kilocalories from fat (approximately .35). The "all meat" subgroups (processed meats and red meat) had positive correlations, while fish and chicken were essentially uncorrelated. Eggs, cheese, and whole milk were also positively correlated ($r = .21$) with percentage of kilocalories from fat, while low-fat milk was negatively correlated ($r = -.19$). The "all breads" food group and starches, representing good sources of complex carbohydrates, showed essentially no correlation to percentage of kilocalories from fat. Fried potatoes alone did, however, show a low positive correlation ($r = .14$). High-fiber bread/cereal and ready-to-eat cereals showed low negative correlations ($r < .20$). Without exception, fruits and vegetables (not including potatoes), as a single food group ($r = -.27$) and when divided into subgroups, were negatively correlated with percentage of kilocalories from fat; the fruit and juice group and the citrus fruit group showed the strongest negative correlations. Alcoholic beverages were inversely correlated with percentage of kilocalories from fat ($r = -.29$). However, the partial correlation between alcoholic beverages and grams of fat (controlling for age, sex, and kilocalories) was $-.50$. As in the case of nutrients, simple correlations were computed for percentage of kilocalories from fat with all food groups by age, sex, race, education, and income subgroups. Of the 432 possible correlations for all food groups by each demographic factor, only 3 were more than $\pm.10$ from the total population simple correlation.

Table 5 shows the median number of servings per week of food groups within each quartile, as well as the ratio of servings per week in the fourth to first quartile in descending order. In comparison with nutrients, the ratios for foods were generally greater. Food groups with the strongest correlation coefficients in Table 4 also tended to show the greatest differences in absolute intake across quartiles. Salty snacks and peanuts had the highest ratios. Absolute servings of table fats were almost twice as great among individuals in the highest than in the lowest quartile. Those in the highest quartile reported eating meat ("all meat" group) more than 10 times per week, as compared with about 6 times per week for those in the first quartile. Results for red and processed meat were consistent with this finding, showing intakes 1.7 and 3.2 times as great in the fourth than in the first quartile, respectively. Servings of fish and chicken in the fourth quartile, however, were 80% of those in the first. Intake of whole milk and cheese was 2.5 times higher in the upper quartile, while low-fat milk intake declined to a median of zero from the first to the fourth quartile. Intake of starches, nonfried potatoes, and all breads showed no major differences across quartiles, but intake of fried potatoes was two times as great in the fourth than in the first quartile. Intake of both ready-to-eat cereal and high-fiber bread/cereal was about half as great in the fourth as in the first quartile. Intakes of fruits and vegetables combined, vegetables excluding potatoes, and fruits alone were all lower among those in the highest quartile. For citrus fruits (primarily oranges) and for fruits and juice, the differences in intake between those in the lowest and highest quartiles were twofold to threefold. Alcoholic beverage intake in the highest quartile was one eighth that in the lowest quartile.

Median intake of food groups and the ratio of quartile 4 to quartile 1 were also determined by race, education, and income categories (results available on request). For salty snacks, peanuts, and low-fat milk, meaningful ratios could not be calculated as a result of zero median intake in quartile 1 or quartile 4. For all meat, all bread, all vegetables, garden vegetables, fruits and vegetables, and citrus fruits, there were no ratio differences greater than $\pm.20$ from the total population ratio presented in Table 5. For the other 18 food groups, 42 of a possible 180 ratios varied from the total population ratio by greater than $\pm.20$, and 22 were greater than $\pm.50$. However, no particular demographic subgroup emerged that had a clear pattern of difference from the total population.

Heavy alcohol consumers tended to be categorized as having a low percentage of kilocalories from fat in their diets. The heaviest drinkers (primarily 18- to 34-year-old men) were easily categorized into what appeared to be the healthiest diet. However, when we either eliminated those consuming more than 20 alcoholic drinks per week from the analyses or removed all kilocalories from alcoholic beverages, establishing new quartile cutpoints for percentage of kilocalories from fat, our overall results regarding either correlations or absolute intake of nutrients or food groups did not change. Therefore, no exclusions were made on the basis of alcohol consumption, allowing for the reality of alcohol intake in the US diet.

Discussion

We have shown, in nationally representative data, that the percentage of kilocalories from fat in the diet is consistently correlated with specific nutrients and food groups across demographic variables of age, sex, race, education, and income, and we found a consistent pattern in the relative change in nutrient and food group intake across quartiles by demographic subgroups. Some of our findings are surprising, while others are not. Given that percentage of kilocalories from fat is different from absolute fat intake, it does not necessarily follow that all foods with substantial fat content will necessarily highly correlate with percentage of kilocalories from fat. The strength of the correlation depends on amount consumed and on fat and caloric content. As expected, high positive correlations were found between percentage of kilocalories from fat and both fat constituents and high-fat food groups. The negative correlations with vitamin C, folate, fiber, carbohydrate, and carotene and associated food groups, however, were less expected. Fruits, vegetables, low-fat milk, and ready-to-eat cereal appear to contribute to a diet low in percentage of kilocalories from fat because they are rich in carbohydrates and substitute carbohydrate kilocalories for fat kilocalories and also because they may be part of a dietary pattern that includes fish and chicken and limits red meats, processed meats, table fats, and eggs. It appears that as individuals decrease their percentage of kilocalories from fat, they substitute primarily carbohydrate- rather than protein-rich foods for energy replacement, partly because protein and fat occur together in many foods.

TABLE 4—Correlations Between Fat and Food Group Intake (Servings per Week): National Health Interview Survey, 1987

Food Group	Simple Correlation[a]	Percentile Correlation[b]	Partial Correlation[c]
Table fats	.37	.38	.37
Processed meats	.36	.41	.39
All meat	.31	.35	.37
Salty snacks	.24	.27	.25
Red meat	.22	.26	.23
Whole milk and cheese	.21	.27	.21
Eggs	.21	.23	.19
Desserts	.17	.20	.18
Fried potatoes	.14	.22	.13
Peanuts	.13	.14	.13
Nonfried potatoes	.05	.07	.05
Dairy products	.03	.04	.02
All bread	−.002	.04	−.03
Stew and soup	−.03	.002	−.02
Starches	−.03	−.01	−.06
Fish and chicken	−.05	−.06	−.05
Salads	−.10	−.09	−.08
All vegetables	−.11	−.09	−.10
Soft drinks	−.12	−.03	−.20
High-fiber bread/cereal	−.15	−.14	−.11
Garden vegetables	−.17	−.15	−.14
Low-fat milk	−.19	−.20	−.17
Ready-to-eat cereal	−.19	−.16	−.16
Fruits and vegetables	−.27	−.25	−.25
Citrus fruit	−.29	−.26	−.27
Fruit and juice	−.29	−.27	−.26
Alcoholic beverages	−.29	−.22	−.50

[a]Pearson correlation of percentage of kilocalories from fat with absolute intake.
[b]Pearson correlation of percentage of kilocalories from fat with age- and sex-specific percentile of intake.
[c]Correlation of grams of fat with absolute intake (controlling for age, sex, and kilocalories).

It is surprising that intakes of nonfried potatoes, starches, and breads changed little with quartiles of percentage of kilocalories from fat since these are foods that Americans are encouraged to consume for purposes of achieving a low-fat diet. In part, this is because of the fats frequently added to these foods (reflected in the NHANES II database used here) that increase their percentage of kilocalories from fat. Furthermore, the stability of these starchy food groups across quartiles suggests that they are staples at all levels of fat intake. Total dairy product intake varied little by fat quartile. However, when the dairy product category is divided into low-fat milk and whole milk and cheese, there is an obvious split in the data indicating that individuals generally consume foods in one group or the other. It is clear that low-fat dairy products can fit easily into a diet low in percentage of kilocalories from fat. Eggs and fish/chicken are both considered medium- to low-fat sources of animal protein, depending on cut and preparation factors. However, in our data, eggs were positively correlated with high-fat diets, while fish/chicken (fried varieties included) were not. These data suggest that eggs are associated with other high-fat eating practices, while fish/chicken consumption characterizes individuals who are not consuming a high-fat diet.

Our analysis focused on the relationship of dietary patterns with fat adjusted for calories, not with absolute fat intake. Because several different techniques are now in use to adjust fat for calories,[22] we used three methods to estimate correlation coefficients with nutrients and food groups. All three methods provided similar results.

Few studies have evaluated data regarding nutrient and food group intake on large representative samples as it relates to percentage of kilocalories from fat. Two small studies, however,

TABLE 5—Median Number of Servings per Week of Foods, by Percentage of Kilocalories from Fat Quartiles: National Health Interview Survey, 1987

Food	Quartile %[a]				Ratio[b]
	1	2	3	4	
Salty snacks	0.2	0.9	1.0	1.0	5.00
Peanuts	0.2	0.5	0.5	0.7	3.50
Processed meats	1.4	2.5	3.5	4.5	3.21
Whole milk and cheese	2.0	3.0	4.0	5.0	2.50
Desserts	2.0	3.2	4.0	4.0	2.00
Eggs	1.0	1.0	2.0	2.0	2.00
Fried potatoes	0.5	1.0	1.0	1.0	2.00
Soft drinks	0.5	1.2	2.0	1.0	2.00
Table fats	7.0	8.0	9.0	12.0	1.71
Red meat	2.3	3.2	3.7	3.9	1.70
All meat	6.2	7.9	9.3	10.4	1.68
All breads	7.0	7.2	7.2	7.2	1.03
Dairy products	7.0	7.0	7.0	7.0	1.00
Starches	5.9	6.1	6.5	5.9	1.00
Nonfried potatoes	2.0	2.0	2.0	2.0	1.00
All vegetables	13.9	13.0	12.9	12.2	0.88
Soup and stew	0.7	0.7	0.7	0.6	0.86
Fish and chicken	2.0	1.9	1.9	1.6	0.80
Salads	6.0	5.1	5.0	4.5	0.75
Garden vegetables	6.2	5.5	5.0	4.5	0.73
Fruits and vegetables	19.8	16.5	14.7	12.5	0.63
High-fiber bread/cereal	4.0	3.0	2.5	2.0	0.50
Fruits and juice	9.2	7.7	6.7	4.4	0.48
Ready-to-eat cereal	3.2	3.0	2.2	1.3	0.41
Citrus fruits	6.7	4.0	3.0	2.0	0.30
Alcoholic beverages	1.7	0.9	0.5	0.2	0.12
Low-fat milk	1.0	0.5	0.0	0.0	0.00

[a]Quartile cutpoints were 33.9%, 38.8%, and 43.3%.
[b]Ratio of quartile 4 to quartile 1.

have investigated changes in total diet as a result of dietary interventions designed to reduce fat intake.[7,8] These intervention studies measured somewhat different effects, since the participants were volunteers motivated by being at high risk for breast cancer and were given detailed instruction in decreasing fat to 15% to 20% of kilocalories while maintaining an adequate overall diet. Furthermore, individuals in our lowest quartile had a median intake of 30% of kilocalories from fat, a value much higher than the mean of approximately 22% achieved in the studies just mentioned.[7,8] Although 4-day food records were used to measure dietary intake in these studies, rather than the food frequency questionnaire approach used here, we nevertheless found similar dietary patterns associated with a low percentage of kilocalories from fat. Absolute mean intake of carbohydrate and protein in the low-fat diets declined very little (10% or less) in both of these studies, while total kilocalorie intake declined by about 25%. Our results show, on the basis of median intake, that individuals with low-fat (30% kilocalories from fat) compared with high-fat diets (46% kilocalories from fat) consume moderately less protein (17%) and energy (14%) but more carbohydrate (22%). In these studies, as in ours, absolute intakes of vitamins A and C[7,8] and folate[8] were higher in the low-fat diets, while calcium, thiamin, riboflavin, and niacin were not greatly affected. With respect to food groups, our data show dietary patterns similar to these studies; subjects with lower relative to higher fat diets reported consistently higher intakes of fruits[7,8] and vegetables,[8] and lower intakes of red and/or processed meat,[7,8] table fats,[7,8] desserts,[7,8] and salty snacks.[8]

Ursin et al.[23] conducted similar analyses of 10 306 adults using a more extensive 93-item food frequency questionnaire administered at the 1982 through 1984 follow-up of the NHANES I cohort. Although this questionnaire and the one used in the present study are different in length, the results for simple correlations between percentage of kilocalories from fat and either nutrient or food group intake are striking in their similarity. Results between the two studies regarding median intake of nutrients and food groups by percentage of kilocalories from fat quartile show that, in general, more marked differences appear in our data moving across quartiles. However, the direction and relative strength of the change for all food groups and nearly all nutrients are the same in both studies. The overall similarities in the results from these two large, nationally based studies provide convincing evidence that certain dietary patterns are consistently associated with a low percentage of kilocalories from fat in the American diet.

One of our initial purposes in conducting this research was to determine differences in correlations and dietary patterns associated with percentage of kilocalories from fat by demographic factors such as age, sex, race, education, and income. Surprisingly, we found few differences. While absolute intake of nutrients and foods did vary by demographics, relative intake, as expressed by our ratios in Tables 3 and 5, varied little by age and sex. Although nutritionists can conceive of many food group intake patterns low or high in percentage of kilocalories from fat, perhaps few distinct patterns actually exist. A possible reason for the few subgroup differences found here is that our short list of food items did not allow variability in dietary patterns to emerge.

In our data, percentage of kilocalories from fat was substantially lower in Hispanics. Our questionnaire, however, has not been compared with dietary records or recalls in this subgroup. In fact, many foods typical for Hispanic subgroups are not included on the questionnaire because such foods do not greatly contribute to total nutrient intake in the US population. In addition, reported intake of nearly all 59 food items was higher among Hispanics than among Blacks and Whites (L. C. Harlan, written communication, January 1993). These considerations, along with the smaller overall numbers of Hispanics, make the data for this important subgroup less trustworthy than those for other racial/ethnic groups, and further research is needed.

Our data show that Americans, on average, continue to consume a diet considered to be too high in fat (mean of 38.8% of kilocalories). Only in the lowest quartile did median values for percentage of kilocalories from fat ap-

proach the recommended level of 30% or less. However, individuals in the lowest relative to the highest quartile appeared to have dietary patterns most consistent with dietary guidelines both from a nutrient and food group perspective. This general finding is consistent with reported intakes of Americans adhering to a low-cholesterol diet in the NHANES II data[24] and among free-living families in Finland.[25] Nutrients of concern (e.g., iron and calcium) frequently found in higher fat foods appeared to be consumed in similar quantities by individuals in each quartile. Although it can be argued that many individuals should be consuming more of these nutrients, the present data suggest that lowering the percentage of kilocalories from fat may not necessarily decrease absolute intake of these important nutrients. As expected, absolute intakes of fatty acids and cholesterol were all lower in the lowest quartile, a generally positive outcome from a public health viewpoint. Vitamin E, however, was also lower, and this nutrient may play a role in both cancer and cardiovascular disease prevention.[26,27] Zinc and magnesium are reported to decline with decreasing fat intake[8]; however, because these minerals were not included on our database, we could not evaluate this possibility. This area of concern should be explored further as low-fat diets are recommended to the population.

It is clear from our data that individuals with a low percentage of kilocalories from fat tend to have dietary patterns notably higher in vitamin C and moderately higher in folate, dietary fiber, and carotene. All of these dietary constituents are being investigated for their role in reducing cancer risk. Our data show that when intakes of fat or nutrients/food groups are investigated, important concurrent dietary factors come into play and must be evaluated. Percentage of kilocalories from fat is highly correlated with certain nutrients and food groups but only weakly associated with others. One cannot be sure, therefore, whether an effect found for lower percentage intakes of kilocalories from fat is really due to higher intakes of vitamin C or folate unless each of these factors is independently evaluated. Assessing total diet is an important factor in sorting out diet and disease relationships.

Another important public health issue involves considering alcoholic beverage intake in studies assessing fat and chronic disease associations. Many heavy drinkers have low intakes of both total fat and percentage of kilocalories from fat and are thus placed in the "healthiest" dietary group, possibly obscuring important findings regarding diet and chronic disease risk. Recommendations for alcohol intake suggest moderation[1-6]; however, from a public health viewpoint, recommending a level of 30% of kilocalories or less for fat still leaves questions regarding how to account for alcohol kilocalories in reducing percentage of kilocalories from fat.

Individuals generally choose to eat foods and not nutrients, and food is selected in the context of a total diet and life-style. We have shown that a single important public health concern, percentage of kilocalories from fat, consistently correlates with selected food groups and nutrients. Individuals with diets low vs high in percentage of kilocalories from fat generally appear to most closely meet current dietary guidelines[6] (an exception is heavy alcohol consumers). The extensive and expensive dietary counseling among motivated women in diet trials[7,8] achieved diets generally lower in percentage of kilocalories from fat than are found in the US population. This approach, however, is not realistic from a public health perspective. Dietary guidelines should advocate a simple food message of increasing intake of fruits, vegetables, whole grains, and low-fat dairy products in a low-fat context. This positive message may not only increase the intake of nutrients/foods that might protect against chronic disease but may also reduce energy intake from fat. □

Acknowledgments

This research was presented at the annual meeting of the Federation of American Societies for Experimental Biology (FASEB), April 1991, Atlanta, Ga.

References

1. *Nutrition and Your Health: Dietary Guidelines for Americans.* 2nd ed. Washington, DC: US Dept of Agriculture; 1985. Home and Garden Bulletin 228.
2. American Cancer Society. Nutrition and cancer: cause and prevention. An American Cancer Society special report. *Cancer.* 1984;34:121–126.
3. *Diet, Nutrition and Cancer Prevention: A Guide to Food Choices.* Washington, DC: National Cancer Institute; 1987. NIH publication 87-2878.
4. *Dietary Guidelines for Healthy American Adults.* Dallas, Tex: American Heart Association; 1986.
5. *The Surgeon General's Report on Nutrition and Health.* Washington, DC: US Dept of Health and Human Services, Public Health Service; 1988. DHHS publication 88-50210.
6. National Research Council, Committee on Diet and Health, Food and Nutrition Board, Commission on Life Sciences. *Diet and Health. Implications for Reducing Chronic Disease Risk.* Washington, DC: National Academy Press; 1989.
7. Gorbach SL, Morrill-LaBrode A, Woods MN, et al. Changes in food patterns during a low-fat dietary intervention in women. *J Am Diet Assoc.* 1990;90:802–809.
8. Buzzard IM, Asp EH, Chlebowski RT, et al. Diet intervention methods to reduce fat intake: nutrient and food group composition of self-selected low-fat diets. *J Am Diet Assoc.* 1990;90:42–50, 53.
9. *The National Health Interview Survey Design, 1973–1984, and Procedures, 1975–1983.* Hyattsville, Md: National Center for Health Statistics; 1985. DHHS publication PHS 85-1320.
10. Block G, Subar AF. Estimates of nutrient intake from a food frequency questionnaire: the 1987 National Health Interview Survey. *J Am Diet Assoc.* 1992;92:969–977.
11. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. *Epidemiology.* 1990; 1:58–64.
12. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. *Am J Epidemiol.* 1986;124:453–469.
13. Block G, Dresser CM, Hartman AM, Carroll MD. Nutrient sources in the American diet: quantitative data from the NHANES II survey. I. Vitamins and minerals. *Am J Epidemiol.* 1985;122:13–26.
14. Block G, Dresser CM, Hartman AM, Carroll MD. Nutrient sources in the American diet: quantitative data from the NHANES II survey. II. Macronutrients and fats. *Am J Epidemiol.* 1985;122:27–40.
15. Sobell J, Block G, Koslowe P, Tobin J, Andres R. Validation of a retrospective questionnaire assessing diet 10–15 years ago. *Am J Epidemiol.* 1989;130:173–187.
16. Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. *J Clin Epidemiol.* 1990;43:1327–1335.
17. Block G, Thompson FE, Hartman AM, Larkin FA, Guire KE. Comparison of two diet questionnaires validated against multiple diet records collected during a 1-year period. *J Am Diet Assoc.* 1992;92: 686–693.
18. Coates RJ, Eley JW, Block G, et al. An evaluation of a food frequency questionnaire for assessing dietary intake of specific carotenoids and vitamin E among low-income Black women. *Am J Epidemiol.* 1991;134:658–671.
19. Harlan LC, Block G. Use of adjustment factors with a brief food frequency questionnaire to obtain nutrient values. *Epidemiology.* 1990;1:224–231.

20. *SAS User's Guide.* Cary, NC: SAS Institute Inc; 1982.
21. *SUDAAN Survey Data Analysis Software.* Research Triangle Park, NC: Research Triangle Institute; 1991.
22. Willett W. *Nutritional Epidemiology.* New York, NY: Oxford University Press; 1990.
23. Ursin G, Ziegler RG, Subar AF, Graubard BI, Haile RW, Hoover R. Dietary patterns associated with a low-fat diet in the National Health Examination Follow-up Study: Identification of potential confounders for epidemiologic analyses. *Am J Epidemiol.* 1993;137:916–927.
24. Schectman G, McKinney WP, Pleuss J, Hoffman RG. Dietary intake of Americans reporting adherence to a low cholesterol diet (NHANES II). *Am J Public Health.* 1990;80:698–703.
25. Pietinen P, Dougherty R, Mutanen M, et al. Dietary intervention study among 30 free-living families in Finland. *J Am Diet Assoc.* 1984;84:313–318.
26. Knekt P. Epidemiological studies of vitamin E and cancer risk. In: Bendich A, Butterworth CE, eds. *Micronutrients in Health and in Disease Prevention.* New York, NY: Marcel Dekker Inc; 1991:142–164.
27. Steiner M. Vitamin E supplementation and platelet function. In: Bendich A, Butterworth CE, eds. *Micronutrients in Health and in Disease Prevention.* New York, NY: Marcel Dekker Inc; 1991:35–68.

Nominations Solicited for Environmental Health Award

The American Academy of Sanitarians is requesting nominations for the 12th annual Davis Calvin Wagner Award. The award, to be presented at the Academy Luncheon during the annual Educational Conference of the National Environmental Health Association, will consist of a plaque and a $500 honorarium. The award is open to all Diplomates of the Academy. Qualifications include the following:

1. The award recipient exhibits resourcefulness and dedication in promoting the improvement of the public's health through application of environmental health and public health practices.
2. The award recipient demonstrates professional, administrative, and technical skill and competence in raising the level of environmental health.
3. The recipient continues to improve him- or herself through involvement in continuing education programs to keep abreast of new developments in environmental health and public health.
4. The recipient is of such excellence as to merit Academy recognition.

Nomination for the award may be made by a colleague or a supervisor and must include the following: name, title, grade, and current place of employment; a description of the nominee's educational background and professional experience; a narrative statement of how the person meets the criteria for the award, including a description of specific accomplishments and contributions on which the nomination is based; and three endorsements (an immediate supervisor and two other members of the professional staff or other persons as appropriate).

The deadline for receipt of nominations is *April 15, 1994.* Three copies of the nomination must be submitted and should be sent to Dr John G. Todd, Chairman, Davis Calvin Wagner Award, 17309 Fletchall St, Poolesville, MD 20837.

The Geographic Variation of Cancer Incidence in Ontario

Stephen D. Walter, PhD, Susan E. Birnie, MBA, Loraine D. Marrett, PhD, S. Martin Taylor, PhD, Donna Reynolds, MSc, John Davies, PhD, John J. Drake, PhD, and Michael Hayes, PhD

ABSTRACT

Objectives. Our objectives were (1) to describe an analysis of the spatial pattern of cancer incidence in Ontario and (2) to discuss the quality of data in the Ontario Cancer Registry with respect to the accuracy of local cancer rates.

Methods. Cancer incidence rates were calculated for 22 cancer sites in 49 counties of Ontario during 1976 to 1986. Capture–recapture methods were used to estimate completeness of case registration, and completeness of residence information was also assessed. Spatial autocorrelation was used in measuring the geographic pattern of incidence rates. Comparisons were also made between sexes and with earlier data from 1966 to 1975.

Results. The quality of the geographic data in the registry appeared good, and corrections for incomplete or inaccurate registration had little impact. About one third of the sex–site combinations showed some evidence of spatial patterning in the cancer rate. Particularly strong regional variation was noted for cancers of the stomach, lung, uterus, and prostate.

Conclusions. The analysis revealed a number of cancers with significant spatial patterning of risk. Further work is needed to relate the cancer data to other information on potential life-style and environmental factors. (*Am J Public Health.* 1994;84:367–376)

Introduction

Cancer is a leading cause of morbidity and mortality in most parts of the world. In Ontario, approximately 34 000 cases are diagnosed annually, and the lifetime cumulative risk of cancer is about 31% for men and 25% for women.[1] One way to investigate the possible impact of environmental factors is through ecologic analysis of spatial and temporal patterns of cancer incidence at the international, national, or regional level.[2–10] Many recent cancer atlases have been used for this purpose,[11] and some studies[12,13] have used formal statistical analyses of the spatial pattern.

The present study examined the pattern of cancer incidence in Ontario to determine whether regions with high cancer risk were spatially aggregated and whether the spatial pattern has persisted over time. The quality of the available data was assessed with respect to the accuracy of local cancer rates.

Methods

Indirectly standardized incidence ratios were calculated for each of 22 cancer sites by sex for each of the 49 provincial counties during 1966 to 1975 and 1976 to 1986, by using Ontario rates as the standard; coefficients of variation over counties were also calculated. Also, the world standard population was used to calculate directly standardized incidence rates for the entire province.

Quality of Data

Cancer incidence data for the periods of study were obtained from the Ontario Cancer Registry, which covers the entire provincial population of 9.1 million (1986 census). Case information was derived from hospital discharges, pathology and regional cancer center reports, and death certificates. Annually, more than 200 000 such records with a mention of cancer are processed. Computerized record linkage and medical logic systems are used to bring together the data on each individual.[14]

Two issues of data quality arise that are relevant to the examination of regional variation in incidence: (1) completeness and accuracy of case registration and (2) completeness and accuracy of residence information. Previous work has shown that accuracy of case registration in Ontario is good, with a high proportion of histologically confirmed cases.[14]

Completeness of Case Registration

The overall completeness of case registration for 1982 was estimated at 95% by means of capture–recapture methodology.[15] The same method was applied here to each site–sex combination to estimate county-level complete-

Stephen D. Walter, Susan E. Birnie, and Donna Reynolds are with the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada. Loraine D. Marrett is with the Department of Preventive Medicine and Biostatistics, University of Toronto, Ontario, and the Ontario Cancer Treatment and Research Foundation, Toronto. S. Martin Taylor, John Davies, and John J. Drake are with the Department of Geography, McMaster University. Michael Hayes is with the Department of Geography, Simon Fraser University, Vancouver, British Columbia, Canada.

Requests for reprints should be sent to Stephen D. Walter, PhD. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre, 1200 Main St W, Hamilton, Ontario, Canada L8N 3Z5.

This paper was accepted June 3, 1993.

TABLE 1—Number of Cases, Age-Adjusted Incidence Rates, and Completeness of Registration and Residence Data, by Cancer Site and Sex: Ontario, 1976 through 1986

Cancer Site/Type (ICD-9 Designation)	Sex	No. of Cases	Age-Adjusted Rate	Coefficient of Variation	Estimated Completeness Rate of Registration, %	Registered Cases Involving Known County of Residence, No. (%)
Lip (140)	M	2 309	4.23	.39	80.8	1 992 (86.3)
	F	405	0.55	.58	70.8	331 (81.7)
Esophagus (150)	M	2 259	4.35	.28	95.6	2 216 (98.1)
	F	1 094	1.53	.38	98.9	1 076 (98.4)
Stomach (151)	M	6 487	11.81	.25	95.9	6 367 (98.2)
	F	4 101	5.54	.32	94.3	3 998 (97.5)
Colon (153)	M	14 888	29.39	.12	93.5	14 456 (97.1)
	F	17 214	25.38	.11	91.6	16 689 (97.0)
Rectum (154)	M	7 975	14.73	.18	95.6	7 702 (96.6)
	F	6 353	9.17	.14	95.7	6 096 (96.0)
Gallbladder (156)	M	1 098	1.98	.32	103.9	1 087 (99.0)
	F	1 761	2.34	.28	98.1	1 741 (98.9)
Pancreas (157)	M	4 815	8.83	.20	97.2	4 776 (99.2)
	F	4 159	5.70	.27	95.2	4 134 (99.4)
Lung (162)	M	34 708	64.84	.17	96.8	34 367 (99.0)
	F	12 260	20.79	.18	96.6	12 497 (99.0)
Bone (170)	M	754	1.38	.73	88.2	734 (97.3)
	F	572	0.97	.49	88.1	563 (98.4)
Melanoma (172)	M	4 084	7.70	.25	90.4	3 699 (90.6)
	F	4 574	7.94	.27	88.7	4 028 (88.1)
Breast (174)	F	41 738	68.72	.01	98.0	40 809 (97.8)
Cervix (180)	F	6 071	10.72	.28	96.0	5 863 (96.6)
Uterus (182)	F	10 643	17.70	.19	94.4	10 253 (96.3)
Ovary (183.0)	F	7 076	11.81	.17	96.1	6 970 (98.5)
Prostate (185)	M	25 020	44.27	.18	94.9	24 005 (95.9)
Bladder (188)	M	12 241	22.24	.18	96.1	11 776 (96.2)
	F	4 308	6.01	.20	94.7	4 098 (95.1)
Kidney (189.0)	M	3 424	6.63	.34	100.2	3 381 (98.7)
	F	2 119	3.52	.22	104.1	2 068 (97.6)
Brain (191, 192)	M	3 820	7.59	.20	97.8	3 788 (99.2)
	F	3 203	5.72	.26	98.3	3 170 (99.0)
Thyroid (193)	M	811	1.54	.49	104.7	786 (96.9)
	F	2 387	4.21	.35	98.4	2 292 (96.0)
Non-Hodgkin's lymphoma (200, 202)	M	5 874	9.03	.20	98.3	5 703 (97.1)
	F	5 201	7.07	.17	96.4	5 060 (97.3)
Hodgkin's disease (201)	M	1 696	3.28	.36	105.8	1 670 (98.5)
	F	1 243	2.27	.38	107.5	1 224 (98.5)
Leukemia (204–208)	M	6 097	11.90	.27	94.7	5 941 (97.4)
	F	4 682	7.62	.43	93.7	4 531 (96.8)

Note. ICD-9 = *International Classification of Diseases*, 9th version. M = male; F = female. Rates are per 100 000.

ness. Log-linear models were fitted[16] with record source types, counties, and their interactions as independent variables. The percentage of missing cases and the corresponding completeness rates were estimated for each county. "Corrected" standardized incidence ratios were calculated after addition of the estimated missing cases to the reported cases.

Comparisons of the uncorrected and corrected standardized incidence ratios indicate the effect (if any) of incompleteness on spatial patterns. The corrected data were adopted unless there was a relatively high negative correlation between the completeness rate and the uncorrected standardized incidence ratio; this might indicate an artificially enhanced spatial pattern in the corrected rates caused by the correction itself.

Completeness of Residence Information

About 3% of cases lack residence information at the county level.[17] There is substantial variation by site, depending largely on the types and numbers of source records contained in the registry, from about 1% for cancers with high hospitalization and case-fatality rates (e.g., lung and pancreas) to about 15% for cancers commonly registered only by pathology reports, which often lack residence information (e.g., melanoma).

In order to use the cases with unknown residence, the geographic referral pattern for each hospital was used to assign residences on a proportional basis. For each hospital, incident cases involving a known residence were used to estimate the proportion of patients from each county. The cases at each hospital involving an unknown residence were then assigned to counties in the same proportions. The final case frequency for each county was derived by summing allocations across hospitals and adding them to the cases involving a known residence.

Accuracy of Residence Information

The concordance of county of residence information within 1 year of diagnosis was examined for patients with multiple records. The coded county of residence, as used in the spatial analysis (normally from the earliest source record after diagnosis), was compared with that from up to five later source records.

Statistical Methods

Spatial analysis of incidence rates. Two methods were used to measure spatial aggregation of the county standardized incidence ratios. Moran's I is a correlation-type index based on continuous data values.[18] The nonparametric rank adjacency statistic D is calculated as the average absolute difference in ranks over pairs of adjacent counties.[19] High values of I and small values of D indicate spatial aggregation (i.e., the tendency for similar rates to be geographically clustered).

Significance levels for I are based on approximate normal z statistics; for D, they must be established through simulation.[19] One-sided significance tests were adopted; two-sided tests are inappropriate since negative autocorrelation would require a "checkerboard" adjacency pattern of high and low rates that would not be plausible for reasonable environmental hypotheses.

Other simulations evaluating the power of these two methods have used various patterns that might result from environmental effects.[20] The power for I was generally high. However, D was also calculated for comparability with previous work elsewhere[12,13]; its power was generally satisfactory, except for certain spatial patterns involving risk elevation in a small number of counties.[20]

I and D were calculated for the completeness rates and the uncorrected and corrected standardized incidence ratios. The ratio I/I_{max} was also calculated, where I_{max} is the maximum attainable value of I in a particular set of data.[18] The computation of I requires a weighting matrix to represent the spatial contact or distance between all possible pairs of counties. In our full set of analyses, a variety of weights was used to represent plausible environmental patterns.[19] For simplicity, we present detailed results only for the simple binary weights indicating adjacency or nonadjacency of county pairs; results with the other weight functions were usually very similar and will be mentioned only in the case of substantial discrepancies.

In determining the strength of clustering in the data, we relied primarily on I as a formal measure of overall spatial aggregation. The mapped data were also visually assessed as a secondary strategy in the identification of high-risk areas. Experimental evidence indicates modest agreement between the visual and statistical methods of data interpretation, although the eye sometimes reacts in unpredictable ways to the way in which data are displayed.[21]

Correlations between sexes, between sites, and over time. The correlations of the county standardized incidence ratios between sexes and between sites within sexes were calculated for the 1976 to 1986 data. Correlations of the county standardized incidence ratios for 1966 to 1975 and 1976 to 1986 for each site and sex were used to examine temporal persistence in spatial patterns; owing to some boundary changes, a few counties had to be combined in this analysis for compatibility over time, resulting in 46 areal units.

Results

The number of cases by site ranged from just over 400 (female lip cancer, rate = 0.55) to 41 880 (breast cancer, rate = 68.7) (Table 1). The coefficients of variation for the uncorrected standardized incidence ratios varied from ratios .01 for breast cancer to .58 for female lip cancer.

Completeness of Case Registration

Table 1 shows that lip and bone cancers and female melanoma had relatively low completeness of registration (<90%). Moderate completeness rates (90% to 95%) were found for colon cancer in both sexes, stomach and uterine cancer in females, and melanoma and prostate and bladder cancer in males. For six sites, the estimated completeness slightly exceeded 100%. This could be due to small sample sizes, dependence between the data sources, or unrecognized duplicate registration. Most sites had very limited geographic variation in completeness of registration.

Completeness of Residence Information

For most cancer sites, more than 95% of registered cases involved a known county of residence (Table 1). For melanoma and lip cancer, which had lower completeness rates, hospital referral patterns were used to make assignments for missing residences. This adjustment had almost no impact on the subsequent spatial analysis. Accordingly, only cases involving a known residence were used for the final analyses of spatial pattern.

Accuracy of Residence Information

The concordance between the residences listed on the diagnosis record and subsequent records was generally high. For example, among patients with at least two records in the registry, the agreement between the diagnosis record and the next source record was 97% for stomach cancer and 94% for lung cancer; even for cancer of the cervix, one of the least reliable sites, the figure was 91%.

Spatial Analysis

General findings. Significance tests based on I and D were usually consistent, and the results found with uncorrected and corrected standardized incidence ratios were similar. Correlations between the uncorrected standardized incidence ratio and the completeness rate were usually small.

No spatial aggregation was found, for either sex, for cancers of the gallbladder, bladder, brain, and thyroid; non-Hodgkin's lymphoma; Hodgkin's disease; or leukemia (Table 2). Also, there was no aggregation for female breast and ovarian cancers. The remaining sites did exhibit spatial aggregation for at least one sex, as discussed below.

Lip cancer. The male data exhibited some spatial aggregation through significant I values and a fairly high I/I_{max}. D was close to significance. There were localized pockets of high and low rates. High rates occurred in certain central and northeastern areas, while rates were low in the urban areas bordering Lake Ontario (Figure 1). Although the completeness rates were relatively low and spatially structured (high completeness rates around Toronto and lower rates in the east and north), their correlation with the standardized incidence ratios was essentially nil. A strong correlation between the two time periods suggests persistence of the spatial pattern.

For females, no strong spatial pattern was evident, but there was a fairly high correlation with the male data. Analyses with some of the alternative weights were significant, but the I/I_{max} values were low. The female spatial pattern may be attenuated by the small numbers of cases, relatively low completeness of registration, and moderate correlation between the uncorrected standardized incidence ratios and completeness rates.

Esophageal cancer. There was weak spatial aggregation in males, with a dispersed pattern of high rates in the north and low rates in some central counties. However, I/I_{max} was not large, and variable significance was found when alternative weights were used. The correlation over time was moderate, but there was almost no correlation between sexes. The negative correlation of completeness rates with the uncorrected standardized incidence ratios may have spuriously enhanced the spatial pattern in the corrected rates. There was no evidence of spatial aggregation in females.

Stomach cancer. Strong spatial aggregation was apparent in both sexes. Northern Ontario and some areas near Lake Ontario had high rates, while central Ontario had lower rates (Figure

TABLE 2—Results of Correlation and Spatial Analyses: Cancer Incidence, Ontario, 1976 through 1986

Cancer Site/Type	Sex	Correlations between			Spatial Statistics		
		SIR and CR	Sexes	Periods	I	I/I_{max}	D
Lip	M	.00	.39	.67	.18*	.42	14.86
	F	.26		−.15	.07	.14	15.75
Esophagus	M	−.37	−.04	.39	.14*	.27	15.08
	F	−.49		.23	−.03	. . .	17.19
Stomach	M	−.03	.61	.71	.30**	.50	13.30**
	F	−.02		.49	.40**	.63	13.15**
Colon	M	.38	.27	.15	−.01	. . .	16.94
	F	.15		.39	.24**	.40	13.25**
Rectum	M	.13	.43	.04	−.04	. . .	15.89
	F	−.39		.36	.15*	.31	13.45**
Gallbladder	M	−.45	.10	.17	.04	.08	15.89
	F	−.13		.26	.05	.11	16.10
Pancreas	M	−.02	.25	.29	.29**	.50	13.40**
	F	.11		.04	−.07	. . .	16.75
Lung	M	.18	.62	.71	.32**	.48	13.44**
	F	.05		.61	.39**	.56	12.51**
Bone	M	.17	.23	.17	−.06	. . .	17.28
	F	−.06		−.21	.16*	.33	15.46
Melanoma	M	.29	.39	.21	.07	.12	15.88
	F	.46		.08	.34**	.45	12.16
Breast	F	.06	. . .	.47	−.01	. . .	17.36
Cervix	F	−.05	. . .	.51	.19*	.32	14.63
Uterus	F	43	. . .	.56	.28**	.40	13.97*
Ovary	F	−.07	. . .	−.01	−.09	. . .	16.62
Prostate	M	.45	. . .	.06	.32**	.49	12.66**
Bladder	M	.19	.22	.22	.09	.25	14.48*
	F	−.21		.22	−.06	. . .	15.96
Kidney	M	−.25	.15	.11	.13*	.21	15.47
	F	−.25		.24	.12	.25	14.63
Brain	M	−.07	.09	−.11	.02	.05	15.50
	F	−.04		−.26	−.10	. . .	17.89
Thyroid	M	−.24	−.03	−.13	.02	.04	16.47
	F	−.35		.35	.08	.15	15.29
Non-Hodgkins lymphoma	M	−.35	.04	−.21	−.02	. . .	16.02
	F	.16		−.09	.06	.10	14.93
Hodgkin's disease	M	−.36	.04	−.02	−.03	. . .	16.10
	F	−.57		.25	.08	.18	15.21
Leukemia	M	.05	.17	.02	.06	.17	17.45
	F	.04		−.03	.02	.06	15.70

Note. SIR = standardized incidence ratio; CR = capture rate; M = male; F = female.
*$P < .05$; **$P < .01$.

2). Significant spatial patterns were found for almost all of the alternative weighting matrices. I/I_{max} values were generally high. Correlations between the sexes and over time were both high.

Colon and rectal cancer. These sites were examined jointly because of the potential for cross classification. However, the correlation of county rates for the two sites was low in both sexes (−.09 in males, −.15 in females), suggesting that systematic errors are unlikely to have influenced the geographic patterns.

Significant spatial patterns were found for both sites in females, with persistence over time. For colon cancer, there were groups of counties with high rates in the northeast and central areas and groups with low rates in the far north and west of Lake Ontario (Figure 3). Most of the alternative weighting matrices for I gave significance, but I/I_{max} values tended to be low. The low geographic variation in incidence may limit the power to detect a spatial pattern for this site.

The spatial aggregation was weaker for rectal cancers in females. Higher rates were more common in the north and near Georgian Bay, and low rates occurred in east-central Ontario and near Lake Erie. Geographic variation in rates was low. Significance in I occurred only with the binary weights. The high negative correlation between the uncorrected standardized incidence ratio and the completeness rate suggests the possibility of an artifact in the spatial pattern of the corrected standardized incidence ratios. However, similar results were found in both the corrected and uncorrected data, indicating a minimal effect of the correction on the analysis.

Pancreatic cancer. Male rates were spatially clustered, with high rates in the north, but no pattern was found for females. Moderate correlation was observed between the sexes, and there was a modest correlation over time in males.

Lung cancer. There were strong geographic patterns for both sexes, with high correlations between sexes and over time. High rates were found in large areas of the northeast and east and in some counties near Lake Ontario and in the southwest (Figure 4). I/I_{max} was generally high.

Bone cancer. While the male data did not demonstrate a spatial pattern, female rates appeared to cluster in small pockets of high and low rates. For females, high rates were found in areas of eastern Ontario and counties bordering Lake Erie. Low rates were found near Lake Ontario. The pattern yielded only a modest value of I, the significance of which did not persist using the alternative weights. There was a weak correlation over time. The small sample size and relatively low completeness of registration may be partly responsible for the inconsistent findings.

Malignant melanoma of the skin. Female rates showed a significant clustering, with a high I/I_{max} value and an elevated risk in a group of counties west and north of Toronto (Figure 5). Rates for males gave some visual impression of a north–south gradient with higher rates in the south, but only one weight matrix for I gave significance. The stronger significance of clustering in females is probably related to the more frequent

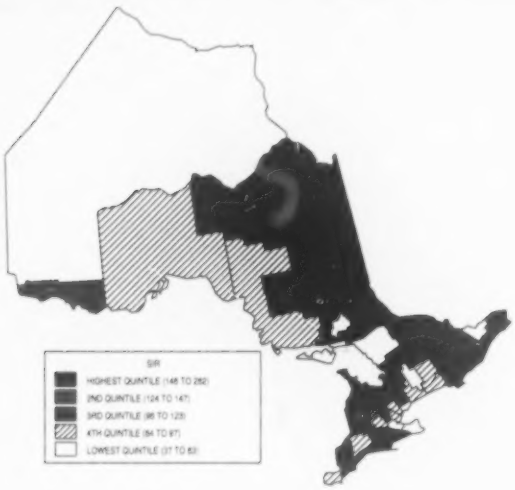

SIR
HIGHEST QUINTILE (146 TO 252)
2ND QUINTILE (124 TO 147)
3RD QUINTILE (96 TO 123)
4TH QUINTILE (64 TO 97)
LOWEST QUINTILE (37 TO 63)

FIGURE 1—Standardized incidence ratios (SIRs) by county, Ontario, 1976 through 1986: male lip cancer.

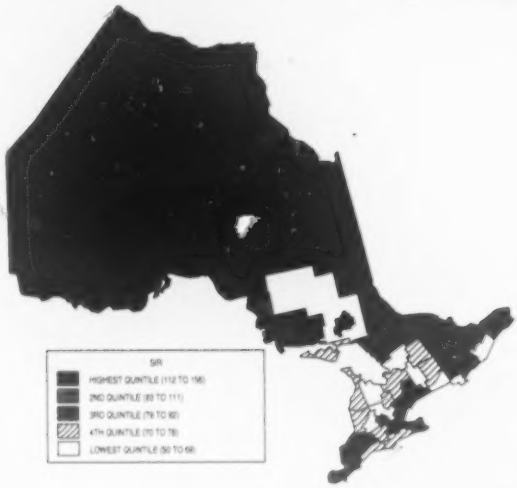


FIGURE 2—Standardized incidence ratios (SIRs) by county, Ontario, 1976 through 1986: male stomach cancer.

contiguity of high-risk counties in the south–central area in comparison with the somewhat less structured pattern for males. Evidence for persistence over time was lacking for both sexes, but the correlation between sexes was relatively high.

Cervical cancer. There was a weak and dispersed geographic pattern with strong persistence over time. High rates tended to occur in northern Ontario and around Georgian Bay, while rates in the southwest were generally low (Figure 6).

Corpus uteri cancer. A significant spatial pattern was seen, with persistence over time. Rates were low in northern Ontario and generally high in the southwest (Figure 7). The effect was dispersed, with large areas of high and low rates. The significant spatial pattern persisted for most of the weighting matrices, and I/I_{max} values were moderately high.

Prostatic cancer. There was strong spatial aggregation, with a high I/I_{max} ratio. A risk gradient by latitude, with low rates in the north and high rates in the southwest, was observed (Figure 8). However, there was a low correlation over time, suggesting a transient or recent effect.

Kidney cancer. The data for males exhibited weak spatial aggregation that did not persist over time. There were small clusters of high and low county rates; however, groups of high rates often adjoined low rates, complicating the overall picture. For females, the spatial pattern was even weaker.

Discussion

Quality of Registry Data

Overall, the quality of the geographic data in the provincial registry was found to be good. The methods used to adjust for incomplete registration and missing residence had little impact on the spatial analyses of the standardized incidence ratios.

Spatial Analysis

Overall, about one third of the sex–site combinations showed some evidence of spatial aggregation, with little difference in results using different weighting schemes. Many of the sites with significant spatial aggregation also evidenced persistence over time (e.g., cancer of the lip and esophagus in males; cancer of the colon, rectum, cervix, and uterus in females; and cancer of the stomach and lung in both sexes). While female thyroid and breast cancers showed no spatial aggregation, they did have a strong correlation over time. It may be that the risk factors for these sites are not spatially aggregated but nevertheless vary geographically.

Interpretation of Significant Spatial Patterns

We now discuss the findings for those cancer sites showing some evidence of spatial aggregation in relation to their major risk factors[22] (summarized in Table 3) and their distributions in Ontario.[23]

Lip cancer. Lip cancer occurs most often in males, among whom significant spatial aggregation and a strong correlation between time periods are apparent. There is also a correlation between sex-specific rates. Lip cancer is strongly associated with exposure to the sun and less so with smoking, particularly of pipes. High rates have been reported among sailors, fishermen, and farmers, presumably because of their solar exposure. In Ontario, the highest rates are found in areas where farming and other outdoor vocations, such as forestry, are common; low rates occur in the urban centers.

Gastrointestinal tract cancers. Several gastrointestinal sites show spatial aggregation in one or both sexes. In addition, there are strong correlations in the male data between stomach and esophageal cancer rates (.48), esophagus and pancreas cancers (.32), and stomach and pancreas cancers (.45).

The most important known risk factors for gastrointestinal cancers are dietary, although the relevant dietary

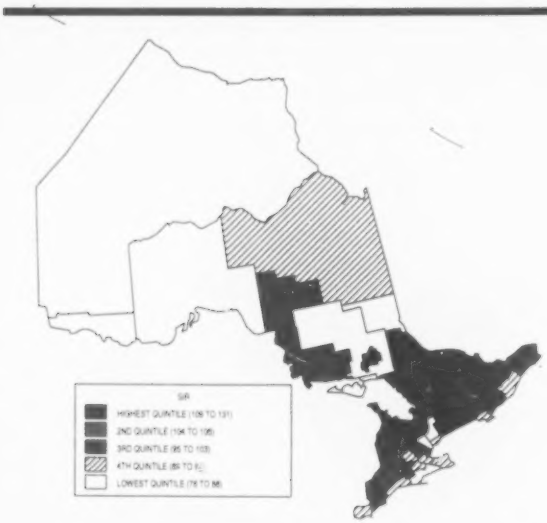


FIGURE 3—Standardized incidence ratios (SIRs) by county, Ontario, 1976 through 1986: female colon cancer.

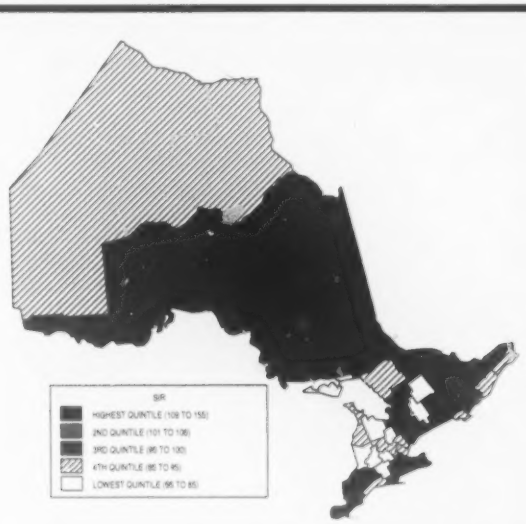

SIR
HIGHEST QUINTILE (109 TO 155)
2ND QUINTILE (101 TO 108)
3RD QUINTILE (95 TO 100)
4TH QUINTILE (85 TO 95)
LOWEST QUINTILE (66 TO 85)

FIGURE 4—Standardized incidence ratios (SIRs) by county, Ontario, 1976 through 1986: male lung cancer.

components vary by site; there are currently no regional data on food consumption patterns in Ontario. Other risk factors include socioeconomic status (SES) (low or high, depending on site), alcohol consumption, and smoking. While there are no regional data on cigarette consumption, there is information on the spatial distribution of SES indicators (e.g., incidence of low income or low education) from census data and of alcohol from regional sales data. It is not known whether SES is directly related to cancer or is a surrogate for some particular aspect of life-style or behavior (e.g., smoking). There is substantial international variation in the incidence of gastrointestinal cancers, and some of the geographic variation found in Ontario may reflect its ethnic diversity.

Esophageal cancer shows spatial clustering in men but not in women. It is associated with consumption of tobacco and alcohol and with low SES. Its pattern in men is generally similar to that for apparent consumption of alcohol. Those who both drink heavily and smoke heavily are at particularly high risk. There is a strong correlation between esophageal and lung cancers in men (.40), and the rate of lung cancer is likely to be a good proxy for smoking prevalence. Consumption of alcohol and tobacco is higher in men, which probably explains both the 3:1 male–female rate ratio and the stronger geographic pattern in men. The distribution of this cancer is similar to that of low education.

Stomach cancer shows the strongest evidence for spatial aggregation and involves strong correlations between sexes and over time. Stomach cancer is particularly interesting because of its consistent temporal decline, its substantial international variation, and its persistence of risk in migrants. Its development is probably related to low intake of fresh fruits and vegetables, high intake of nitrates and nitrites, and low SES. Its pattern in Ontario is similar to that for low education. The strong correlations of stomach cancer with esophageal cancer in males (.48) and with lung cancer in both sexes (.56 and .34 in males and females, respectively) may partly reflect the fact that all of these cancers are more common in people of lower SES. Likewise, a negative correlation with colon cancer in females (−.33) may reflect colon cancer's association with higher SES. The fact that some of these correlations are sex specific implies that SES may be an indicator of different factors in the different cancer sites (e.g., for stomach cancer, it may be a surrogate for a high-risk diet; for lung cancer, it may indicate smoking status).

Cancers of the colon and rectum cluster in females only, although there is a moderate correlation between sexes and persistence over time in females only. These two cancers are not correlated with each other; however, there is a moderate correlation between rectal cancer and lung cancer in both sexes (.28 and .30 in males and females, respectively). Colo-rectal cancers are believed to be related to high levels of fat and low levels of fiber consumption, although the data are less consistent for rectal cancer. Rectal cancer may also be related to intake of alcohol, and its pattern appears similar to that for alcohol sales in Ontario. The pattern of colon cancer is roughly inverse to that for low education; conversely, rectal cancer is distributed similarly to low levels of education.

Cancer of the pancreas, which appears to cluster in males but not females, is associated with cigarette smoking. In males, there is a moderate correlation between rates for pancreas and lung cancers (.27), which suggests that geographic aggregation in pancreas cancer may be related to variation in smoking rates. Pancreatic cancer may also be related to diet, but studies have produced conflicting results regarding particular dietary constituents. Its correlations with esophagus (.32) and stomach cancer (.45) in males suggest the possibility of common risk factors.

Lung cancer. Lung cancer shows strong spatial aggregation in both sexes, with a strong correlation between sexes and over time. It is well known to be strongly associated with tobacco con-

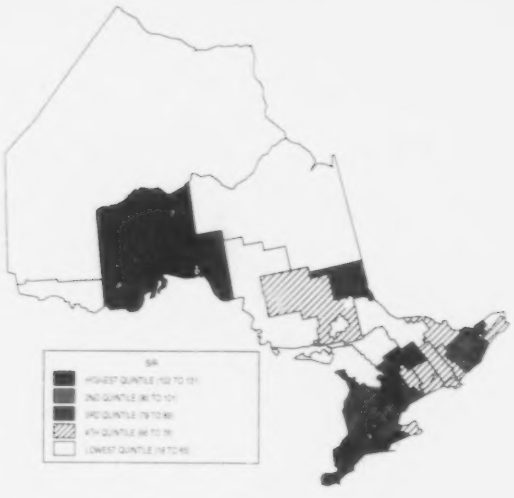


FIGURE 5—Standardized incidence ratios (SIRs) by county, Ontario, 1976 through 1986: female melanoma.

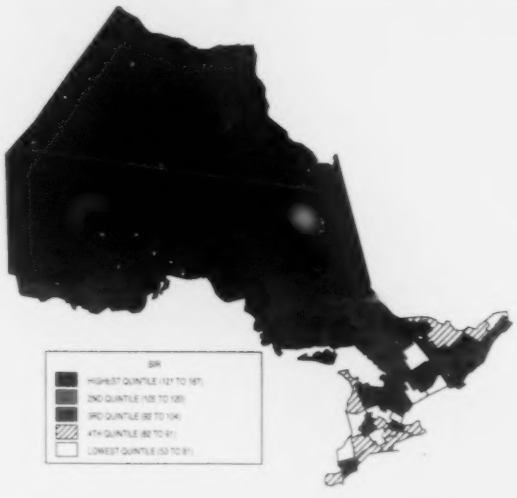


FIGURE 6—Standardized incidence ratios (SIRs) by county, Ontario, 1976 through 1986: cervical cancer.

sumption, but it is also more common in lower SES groups, and some cases are associated with occupational exposures to inhaled carcinogens such as asbestos, radon, and nickel dust. An analysis of Ontario lung cancer mortality found positive associations with employment in mining and prevalence of low education and a negative association with agricultural activity.[24]

Bone cancer. Bone cancer shows aggregation in females only, with a negative correlation between time periods. Although this is a relatively rare cancer, it is one of the most common tumors of adolescence. Except for ionizing radiation, little is known about its risk factors. It is interesting that one cluster of high rates occurs in the specialized agricultural areas bordering Lake Erie and the southern shore of Lake Ontario, where tender fruits and tobacco are grown.

Malignant melanoma of the skin. Melanoma shows a strong spatial pattern in females, with a moderate correlation between the sexes and no persistence over time. Its incidence is correlated with cancers of the corpus uteri (.56) and prostate (.44), possibly through common socioeconomic effects. Its incidence has been increasing rapidly in Ontario and in other light-skinned populations. Major risk factors include exposure to ultraviolet radiation, fair pigmentation, and high SES. Rates are higher in the southern part of the province, where solar ultraviolet radiation is most intense. However, the areas with the highest rates are those where cottaging or farming are common and exposure to solar ultraviolet radiation is presumably high. Melanoma rates show a weaker relationship to farming areas in males. In general, groups of people with high cumulative ultraviolet exposure (such as farmers) have a lower risk of melanoma than indoor workers. The male–female difference might therefore be related to some protective effect afforded by a higher chronic exposure level in males. Areas with the highest rates of melanoma are also those with the greatest proportions of people of British or German ethnicity, who are likely to be of fairer complexion. A recent study in Ontario indicated a risk associated with the use of sunbeds and sunlamps,[25] but the contribution of this factor to the geographic distribution is not known.

Genitourinary cancers. Cancers of the cervix and corpus uteri both show spatial clustering. The two sites have a moderate negative correlation (−.21), and their geographic patterns appear to be very different. As would be expected from its known associations with smoking, cervical cancer is positively correlated with lung (.32) and bladder (.34) cancers. In addition, its pattern is similar to that for prevalence of low education. There are no regional data available concerning sexual factors (e.g., age at first coitus or prevalence of venereal agents such as human papilloma viruses). Cervical cancer incidence has been declining in Ontario, at least in part because of screening. It is likely that there is geographic variation in screening intensity, although no such data are currently available.

Risk factors for uterine and cervical cancers are very different. Corpus uteri cancer is a hormone-dependent tumor resulting from high levels of unopposed estrogenic stimulation. It is primarily a disease of postmenopausal women who either are obese or have used replacement estrogen therapy during or after menopause. Nulliparous women are also at increased risk. Use of combination oral contraceptives, which guarantee exposure to progesterone along with estrogen, appears to reduce risk. In Ontario, the incidence of uterine cancer is correlated with that for melanoma and breast cancer, two other cancers associated with higher SES. There are no data available to examine regional variation in rates of obesity, nulliparity, or estrogen use, but the distribution of this cancer is roughly inverse to that for low education and low income.

The prostate is the second most common cancer site in males (after lung cancer), and it is strongly spatially aggregated. However, there is no evi-

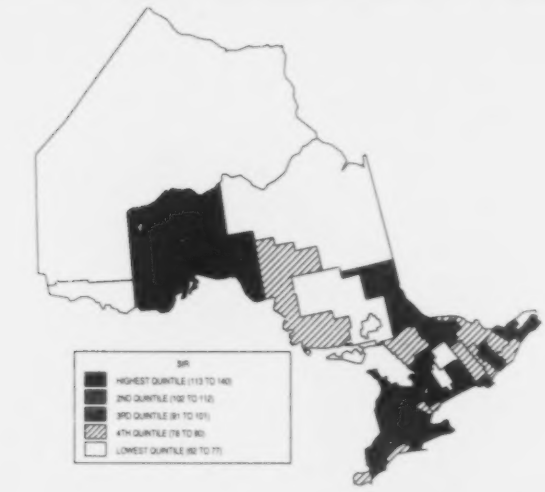

SIR
HIGHEST QUINTILE (113 TO 140)
2ND QUINTILE (102 TO 112)
3RD QUINTILE (91 TO 101)
4TH QUINTILE (78 TO 90)
LOWEST QUINTILE (62 TO 77)

FIGURE 7—Standardized incidence ratios (SIRs) by county, Ontario, 1976 through 1986: uterine cancer.

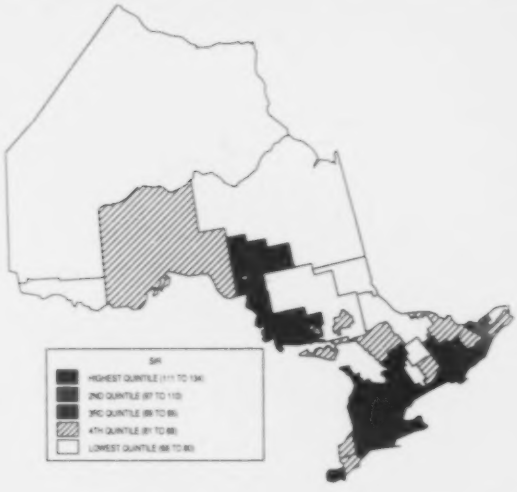


FIGURE 8—Standardized incidence ratios (SIRs) by county, Ontario, 1976 through 1986: prostate cancer.

TABLE 3—Major Known Risk Factors for Cancer Sites/Types with Evidence of Spatial Aggregation

Site/Type	Risk Factor(s)
Lip	Sunlight, tobacco
Esophagus	Tobacco, alcohol, diet low in fresh fruits and vegetables, low socioeconomic status
Stomach	Diet low in fresh fruits and vegetables, high intake of nitrates/nitrites (e.g., in water and preserved foods), low socioeconomic status
Colon/rectum	Diet high in fat, diet low in the fiber, high socioeconomic status (colon)
Pancreas	Tobacco
Lung	Tobacco, occupational exposures (radon, asbestos, etc.), low socioeconomic status
Bone	Ionizing radiation
Malignant melanoma	Ultraviolet radiation, susceptible phenotype (fair hair/eyes/skin), high socioeconomic status
Cervix uteri	Tobacco, sexual promiscuity (early age at coitus, multiple sex partners), viruses (human papilloma virus, herpes simplex virus type 2), low socioeconomic status
Corpus uteri	High levels of unopposed endogenous estrogens, use of exogenous estrogens (unopposed), use of combination oral contraceptives (protective factor), high socioeconomic status
Prostate	Diet (possibly high in fat), high levels of androgens
Kidney	Tobacco, obesity

dence of persistence of the spatial distribution over time. In our data, prostate cancer incidence is negatively correlated with esophageal and stomach cancers (−.47 and −.30, respectively), again possibly as a result of socioeconomic effects. There is substantial racial and international variation in the rate of prostatic cancer: US Blacks have the highest incidence, while Asian populations have very low rates. Upon migration to the United States, however, Asians' rates increase.

Little is known about the etiology of this cancer; the effects of dietary fat and male hormones are still speculative. Smoking may slightly increase risk, perhaps through its effect on male hormones.[26] There is a known high prevalence of latent tumors in asymptomatic older men. These tumors are sometimes detected during transurethral resection of the prostate for the treatment of benign prostate disease, which may partly explain why incidence is increasing but mortality is not.

In Ontario, the spatial pattern appears to be inverse to that for low education and is dissimilar to the pattern observed for mortality rates.[23] Thus, regional variation in disease detection rates may be a determinant of the incidence pattern. Examination of the incidence of this tumor in men under the age of 65, who would be less likely to receive transurethral resection of the prostate or general screening, may be informative. Completeness rates for registration of prostatic cancer were significantly spatially aggregated and were positively correlated with the uncorrected standardized incidence ratios; this would tend to have obscured a spatial pattern in the corrected standard-

ized incidence ratios, so the apparently strong pattern is not likely to be an artifact of variation in registration rates.

Kidney cancer shows weak clustering in males. Its incidence is correlated with that of esophageal, stomach, pancreas, and lung cancers, particularly in males. All of these sites except the stomach have been consistently associated with tobacco consumption (the strongest known risk factor), and most are common in lower SES groups. While obesity increases risk, the role of fat consumption and use of certain medications is uncertain.

Implications for Future Study

Several cancer sites have been shown to have significant regional variation in Ontario. The explanation for the variability, however, requires further study, especially concerning the possible impact of environmental factors. In related research to create an inventory of existing data sources in the province, a provincewide health survey and seven atmospheric, five hydrologic, one geologic, four land-use, two sociodemographic, and nine industrialization databases were identified as being potentially useful for ecologic studies.[27] The health survey, in particular, will provide regional data on many risk factors of interest: the 1990 Ontario Health Survey used a regionally stratified sample of about 47 000 Ontario residents and included questions concerning smoking, health behavior, diet, beverage consumption, occupation, and socioeconomic factors. Previous work on cancer mortality in Ontario[24] has revealed substantial variation between counties with respect to ethnic, industrial, and SES characteristics in censal data.

Ecologic studies of the kind described here provide a relatively inexpensive way of examining regional variation in health in large populations. The effects of environmental factors can also often be addressed by access to existing data sets.[28,29] However, these studies involve interpretational problems arising from the aggregated form of the data and effects such as variation in the size of the regional populations, migration, and disease latency.[19,28,29]

The census divisions of Ontario are quite variable in population size and area. However, the regional data values used in our spatial analyses were all given equal weight, so one needs to be less concerned about the possibly dominant effects of the large cities (here, notably, Toronto). Isodemographic map projections have been proposed to eliminate problems associated with heterogeneous population density. We decided against this option when mapping our data because the Ontario map becomes physically unrecognizable when this projection method is used; furthermore, one loses the information on the relative distance between regions, and physical distance is a more important determinant of interregional correlations than is population density.

Migration and latency are more difficult problems. The exposures relevant to cancer risk probably occur 10 to 20 years before diagnosis; thus, current residence will be an imperfect marker. Our use of incidence data eliminates any potential bias associated with postdiagnostic migration. Any bias due to prediagnostic migration of future cancer cases would most likely be evidenced by a relative cancer excess in urban areas near major treatment centers. Census data[30] indicate that 16% of the Ontario population in 1986 had moved from another census division since 1981, with relatively small differences in this rate across the province; this suggests that migration bias is unimportant. Furthermore, the spatial patterns observed in the incidence data are not compatible with an effect of this kind. We conclude that migration in the Ontario population is more likely to attenuate an environmental effect than to create bias.

With its large population, interesting regional patterns of cancer incidence, and known variation in risk factors for some cancers, Ontario appears to be a suitable location for further study of the relationship between environment and cancer. Future research might use the ecologic approach to relate cancer incidence and relevant "exposures" or adopt case–control and other designs with individually based data. □

Acknowledgments

This study was supported by a grant from the Ontario Ministry of Health. S. D. Walter holds a National Health Scientist Award from the National Health Research and Development Program, Health and Welfare Canada.

We would like to acknowledge the assistance of A. Chiarelli, G. Fehringer, M. Swift, P. Toogood, and S. Wort in various aspects of executing this study.

References

1. *Cancer in Ontario 1988.* Toronto, Ontario, Canada: Ontario Cancer Treatment and Research Foundation; 1988.
2. Blot WJ, Fraumeni JF. Geographic epidemiology of cancer in the United States. In: Schottenfeld D, Fraumeni JF, eds. *Cancer Epidemiology and Prevention.* Philadelphia, Pa: WB Saunders; 1982;10:179–193.
3. Savitz DA, Redmond CK. Screening for geographic heterogeneity of disease rates: application to cancer incidence in Allegheny County, Pennsylvania, 1969–71. *J Chronic Dis.* 1985;38:145–156.
4. Hoover R. Geographic patterns of cancer mortality in the United States. In: Fraumeni JF, ed. *Persons at High Risk of Cancer.* San Diego, Calif: Academic Press; 1975;20:343–360.
5. Kendrick BL. A spatial, environmental and socioeconomic appraisal of cancer in New Zealand. *Soc Sci Med.* 1980;14D:205–214.
6. Greenberg MR, Preuss PW, Anderson R. Clues for case-control studies of cancer in the northeast urban corridor. *Soc Sci Med.* 1980;14D:37–43.
7. Minowa M, Shigematsu I, Nagai M, et al. Geographical distribution of lung cancer mortality and environmental factors in Japan. *Soc Sci Med.* 1981;15D:225–231.
8. Najem GR, Thind IS, Lavenhar MA, et al. Gastrointestinal cancer mortality in New Jersey counties, and the relationship with environmental variables. *Int J Epidemiol.* 1983;12:276–288.
9. Wigle D, Mao Y, Semenciw R, et al. Cancer patterns in Canada. *Can Med Assoc J.* 1986;34:231–235.
10. Rose G. Causes of the trends and variations in CHD mortality in different countries. *Int J Epidemiol.* 1989;18:S174–S179.
11. Walter SD, Birnie SE. Mapping mortality and morbidity patterns: an international comparison. *Int J Epidemiol.* 1991;20:678–689.
12. *Atlas of Cancer in Scotland: 1975–1980.* Lyon, France: International Agency for Research on Cancer; 1985. Scientific Publication 72.
13. Cislaghi C, Decarli A, La Vecchia C, Laverda N, Mezzanotte G, Smans M. *Statistics and Maps on Cancer Mortality, Italy, 1975–1977.* Bologna, Italy: Pitagora Editrice; 1986.
14. Clarke EA, Marrett LD, Kreiger N. *Twenty Years of Cancer Incidence 1964–1983: Ontario Cancer Registry.* Toronto, Ontario, Canada: Ontario Cancer Treatment and Research Foundation; 1987.
15. Robles SC, Marrett LD, Clarke EA, Risch HA. An application of capture-recapture models to the estimation of completeness of cancer registration. *J Clin Epidemiol.* 1988;41:495–501.
16. *GLIM. Generalised Linear Interactive Modelling, Version 3.77.* London, England: Royal Statistical Society; 1978.
17. *Completeness of Residence Information.* Ontario, Canada: Ontario Cancer Registry; 1988.
18. Cliff AD, Ord JK. *Spatial Processes:*

Models and Applications. London, England: Pion; 1981.

19. Walter SD. The analysis of regional patterns in health data: I. Distributional considerations. *Am J Epidemiol.* 1992;136: 730–741.
20. Walter SD. The analysis of regional patterns in health data: II. The power to detect environmental effects. *Am J Epidemiol.* 1992;136:742–759.
21. Walter SD. Visual and statistical assessment of spatial clustering in mapped data. *Statistics Med.* In press.
22. Tomatis L, ed. *Cancer: Causes, Occurrence, and Control.* Lyon, France: International Agency for Research on Cancer; 1990. Scientific publication 100.
23. Marrett LD, Swift MB, Reynolds DL, Clarke EA. *Geographic Distribution of Cancer in Ontario. Vol 1: Atlas of Cancer Mortality, 1976–1985.* Toronto, Ontario, Canada: Ontario Cancer Treatment and Research Foundation; 1991.
24. Reynolds DL. *The Geographic Distribution of Cancer Mortality in Ontario: An Ecologic Analysis on Ethnicity, Industry and Socioeconomic Characteristics, 1976–1985.* Toronto, Ontario, Canada: University of Toronto; 1989. Thesis.
25. Walter SD, Marrett LD, From L, et al. The association of cutaneous malignant melanoma with the use of sunbeds and sunlamps. *Am J Epidemiol.* 1990;131:232–243.
26. Hsing AW, McLaughlin JK, Schuman LM, et al. Diet, tobacco use, and fatal prostate cancer: results from the Lutheran brotherhood cohort study. *Cancer Res.* 1990;50:6836–6840.
27. Walter SD, Marrett LD, Taylor SM, et al. *Geographic Variation in Cancer Rates.* Toronto, Ontario, Canada: Ontario Ministry of Health; 1991.
28. Walter SD. The ecologic method in the study of environmental health: I. Overview of the method. *Environ Health Perspect.* 1991;94:61–65.
29. Walter SD. The ecologic method in the study of environmental health: II. Methodologic issues and feasibility. *Environ Health Perspect.* 1991;94:67–73.
30. Population and dwelling characteristics—census divisions and subdivisions; Ontario, part 2. In: *Profiles Census Canada 986.* Ottawa, Ontario, Canada: Government of Canada; 1988. Publication no. 94–112.

Using CD4 Counts to Evaluate the Stages and Epidemiology of HIV Infection in South Carolina Public Clinic Patients

Stephen Luby, MD, Jeffrey Jones, MD, MPH, and John Horan, MD, MPH

ABSTRACT

Objectives. CD4 lymphocyte counts decrease with the duration of human immunodeficiency virus (HIV) infection. We used CD4 counts collected for clinical reasons to evaluate the stage of HIV infection and the epidemiology of recent HIV infections among attendees of South Carolina's public health clinics.

Methods. We measured the CD4 T-lymphocyte counts of persons newly diagnosed with HIV infection April 1989 through June 1990 at South Carolina public health clinics who returned for follow-up.

Results. Of 812 newly diagnosed HIV-infected health department patients, 420 (52%) had their CD4 lymphocyte counts measured. Of these 420, 51 (12%) had CD4 counts of <200, the level below which prophylaxis for pneumocystis pneumonia prolongs survival, and 193 (46%) had CD4 counts of <500, the level below which zidovudine may prolong disease-free survival. The highest CD4 counts (≥ 900), which are associated with more recent HIV infection, were more common in females.

Conclusions. In South Carolina, almost half of newly reported HIV-infected persons who agreed to CD4 testing at the health department might benefit from immediate drug therapy. Within this population, women may be an emerging risk group that requires specifically directed HIV prevention efforts. (*Am J Public Health.* 1994;84:377–381)

Introduction

Several cohort studies demonstrated that CD4 lymphocyte counts decrease with the duration of human immunodeficiency virus (HIV) infection.[1–7] This decrease tends to be linear[8] and predicts opportunistic infection and progression to the acquired immunodeficiency syndrome (AIDS).[2–9] Thus, measuring the CD4 lymphocyte count of HIV-infected persons helps characterize the HIV/AIDS epidemic by staging HIV infection in a population. Indeed, CD4 counts could supplement HIV surveillance by characterizing groups with advanced HIV disease in need of drug therapy as well as identifying those persons most recently infected to evaluate changing demographics and risk groups for HIV transmission.

However, previous studies of CD4 counts in HIV-infected persons have been from restricted populations, including cohorts of homosexual men,[1,3,6,10,11] hemophiliacs,[2,12] attendees at two sexually transmitted disease clinics,[13] and members of the US Army,[14] which is 91% male and attempts to exclude persons who are homosexual, intravenous drug users, or HIV infected.

In March 1989, the South Carolina Department of Health and Environmental Control began offering CD4 lymphocyte testing free of charge at 81 clinics statewide to all HIV-infected persons reported to the health department. South Carolina law requires that all physicians, hospitals, and laboratories licensed by the state report persons with HIV infection to the Department of Health and Environmental Control.

Clinic nurses use these CD4 counts to identify persons who are late in the course of their HIV disease and therefore need immediate medical services. Persons with CD4 counts of <500 who are indigent qualify for Medicaid and are eligible to receive zidovudine through the federal drug reimbursement program. Persons with CD4 counts of <200 receive urgent medical referral to receive zidovudine and prophylaxis against *Pneumocystis carinii* pneumonia.

For this study, we retrospectively reviewed the CD4 lymphocyte counts of newly identified HIV-infected persons in the first 15 months of the South Carolina CD4 testing program. Our objectives were to determine the percentage and characteristics of HIV-infected persons who might benefit from drug therapy and to characterize the population of most recently infected persons.

Methods

Patients

We selected patients newly reported to the state's HIV registry April 1989 through June 1990 whose CD4 T-lymphocyte count was measured in the 90 days after their first positive HIV serology test. Hence, CD4 tests were

Stephen Luby and John Horan are with the Division of Field Epidemiology, Centers for Disease Control and Prevention, Atlanta, Ga. At the time of this study, Jeffrey Jones was with the South Carolina Department of Health and Environmental Control, Columbia, SC; he is currently with the Division of HIV/AIDS, Centers for Disease Control and Prevention, Atlanta, Ga.

Requests for reprints should be sent to Jeffrey Jones, MD, MPH, Division of HIV/AIDS, MS E-47, Centers for Disease Control and Prevention, Atlanta, GA 30333.

This paper was accepted June 3, 1993.

Note. Inclusion of trade names is for identification only and does not imply endorsement by the Department of Health and Human Services or the US Public Health Service.

TABLE 1—Rates of CD4 Testing of HIV-Infected State Clinic Patients, South Carolina, April 1989 through September 1990

	Total	No. (%) CD4 Tested	Relative Rate	95% CI	P
Sex					
Female	223	126 (57)	1.1	1.0, 1.3	.11
Male	589	294 (50)			
Location					
Rural	490	259 (53)	1.1	0.9, 1.2	.47
Urban	322	161 (50)			
Race[a]					
White	186	121 (65)	1.4	1.2, 1.6	<.001
Black	618	296 (48)			
HIV risk factor					
None reported	61	18 (30)	0.6	0.4, 0.8	<.001
Heterosexual contact	412	216 (52)			
Male-to-male sex	245	141 (58)	χ^2 association; $P = .50$		
Intravenous drugs	85	41 (48)			
Transfusion	7	3 (43)			
Hemophilia	2	1 (50)			

[a]Eight persons who reported their race as Hispanic are not included in the racial breakdown.

performed April 1989 through September 1990. A 90-day time frame was chosen to permit timely analysis of the data. We excluded HIV-infected children younger than 13 years old because their absolute CD4 counts are not directly comparable with adult CD4 counts.[15,16]

Specimen Handling

Nurses in the health department clinics collected blood samples from consenting HIV-infected patients. For patients originally identified in these health department clinics, CD4 lymphocyte counts were measured at the first posttest counseling session: that is, when the patient first learned of his or her HIV infection. All blood samples were analyzed by flow cytometry within 20 hours of being drawn with a Coulter Epics–Profile II flow cytometer (Coulter Corporation, Hialeah, Fl). Personnel in the South Carolina Department of Health and Environmental Control Bureau of Laboratories entered the absolute CD4 lymphocyte count (the product of the total lymphocyte count and the percentage of CD4+ lymphocytes) with the patient's name onto the state's mainframe computer, and the result was sent back to the originating clinic.

Data Analysis

The CD4 tests were matched by name, date of birth, county, and race with the Department of Health and Environmental Control surveillance registry of HIV-infected persons by one person working within the HIV/AIDS surveillance section of the department. For analysis we compiled a linked CD4 count/HIV registry microcomputer file with all personal identifiers removed.

We classified persons with absolute CD4 counts of ≥ 900 as recently infected and persons with CD4 counts of < 200 as having advanced HIV infection. We chose a count of 900 as a threshold for defining recent infection because previous researchers found 918 to be the average count for six uninfected control persons who were repeatedly tested.[17] A larger study has since found an average CD4 count of 890 among 63 repeatedly tested HIV-negative control subjects.[18] We chose CD4 counts of < 200 to indicate advanced infection because at this level persons have markedly increased risk of opportunistic infection.[9]

We initially evaluated age groups at the standard intervals used for the surveillance of HIV/AIDS—that is, ages 5–12 years, 13–19 years, and at 5-year intervals thereafter.[19] Because children under age 13 years were not analyzed, persons aged 13–19 comprised the youngest group analyzed. Because only 5% of the study population was in this youngest age group, we expanded the youngest category to include persons 20–24 years old to permit a more thorough statistical evaluation of age.

We calculated relative risks (RRs) and confidence intervals (CIs) by the method of Greenland and Robins using Epi Info software.[20] We used StatXact[21] software to perform significance testing of dichotomous outcomes by Fisher's exact chi-square association. Using Kwikstat software,[22] we compared mean ages across two groups by Student's *t* test and assessed the association between age and CD4 cell count using Pearson's correlation.

Results

Of the 1510 newly diagnosed HIV-infected persons more than 12 years old reported to the South Carolina Department of Health and Environmental Control April 1989 through June 1990, 812 (54%) were initially diagnosed as HIV infected at state clinics and 698 (46%) were reported from outside the state clinic system. Persons identified outside the state clinic system were significantly less likely to participate in the CD4 lymphocyte testing program compared with state clinic–identified patients (20% vs 52%; RR = 0.39, 95% CI = 0.33, 0.46) and were 1.9 times more likely to have CD4 counts of < 200 (23% vs 12%; RR = 1.9, 95% CI = 1.3, 2.9). Because persons diagnosed outside the clinic system were less completely represented in the database and because those non–public clinic patients who were tested appeared to overrepresent a selected group of persons with more advanced disease, we limited subsequent analysis to the 812 state clinic–identified patients.

Of these 812 newly reported HIV-infected patients, 420 (52%) had their CD4 lymphocyte counts measured at a state clinic within 90 days of their first positive HIV serology test. The mean age of clinic patients whose CD4 counts were measured was 30.7 years, which was similar to the mean of 30.2 years among HIV-infected clinic patients whose CD4 counts were not measured. Gender and urban or rural residence were not significantly associated with undergoing CD4 testing (Table 1). Whites were 1.4 times more likely than Blacks to participate in the lymphocyte testing program. Persons who reported no HIV risk factors were significantly less likely to undergo follow-up CD4 testing. No defined risk group for acquiring HIV infection was significantly more or less likely to be CD4 tested.

Within the study population of 420 newly diagnosed HIV-infected persons

TABLE 2—Distribution of CD4 Counts of HIV-Infected State Clinic Patients, South Carolina, April 1989 through September 1990

	Total	No. (%) with CD4 of <200	No. (%) with CD4 of >199 and <500	No. (%) with CD4 of >499 and <900	No. (%) with CD4 of ≥900	P (χ^2)
All patients	420	51 (12)	142 (34)	158 (37)	68 (16)	
Age, y						
>24	322	50 (16)	109 (34)	109 (34)	54 (17)	<.01
≤24	98	1 (1)	33 (34)	49 (50)	15 (15)	
Sex						
Female	126	10 (8)	34 (27)	53 (42)	29 (23)	.014
Male	294	41 (14)	108 (37)	105 (36)	40 (14)	
Race[a]						
Black	299	38 (13)	99 (33)	112 (37)	50 (17)	.871
White	119	12 (10)	43 (36)	45 (38)	19 (16)	
Location						
Rural	259	35 (14)	78 (30)	105 (41)	41 (16)	.145
Urban	161	16 (10)	64 (40)	53 (33)	28 (17)	
HIV risk factor						
None reported	18	5 (28)	4 (22)	4 (22)	5 (28)	
Heterosexual contact	216	24 (31)	73 (34)	83 (38)	36 (17)	
Male-to-male sex	141	16 (11)	52 (37)	55 (39)	18 (13)	
Intravenous drugs	41	6 (15)	11 (27)	15 (37)	9 (22)	.520
Transfusion	3	0 (0)	1 (33)	1 (33)	1 (33)	
Hemophilia	1	0 (0)	1 (0)	0 (0)	0 (0)	

[a]Two Hispanic persons are not included in the racial breakdown. Totals for racial groups differ slightly from those reported in Table 1 because five persons were recorded as a different race when they were HIV tested than they were during subsequent CD4 testing.

from state clinics, 51 (12%) had what we defined as advanced HIV disease (CD4 counts of <200). One hundred ninety-three persons (46%) had CD4 counts of <500. At the other end of the CD4 spectrum, 69 persons (16%) had CD4 counts of ≥900. The mean CD4 count for the population was 588 with a standard deviation of 347.

The distribution of CD4 counts by age, sex, race, location, and HIV risk factor is shown in Table 2. Females were 1.7 times more likely than males to have CD4 counts of ≥900 (95% CI = 1.1, 2.6). No significantly increased risks for recent infection were associated with age, race, urban versus rural residence, or risk group for acquiring HIV infection.

Reported risk factors for acquiring HIV infection for the 29 females who had CD4 counts of ≥900 are listed in Table 3. Among these 29 women, 21 (72%) reported heterosexual contact as their primary risk behavior for acquiring HIV infection. The proportional distribution of reported risk factors is not significantly different between females with CD4 counts of ≥900 and those with CD4 counts of <900.

At the other extreme of CD4 counts, persons more than 24 years old had a 15.2 times greater prevalence of advanced HIV disease than persons 24 years of age or younger (95% CI = 2.1, 108.7). Males were 1.8 times more likely than females to have advanced infection (95% CI = 0.9, 3.4). Persons who did not report an HIV risk factor were 2.4 times more likely to have CD4 counts of <200, but there were no other significant differences by HIV risk group, race, or urban versus rural location. There was no statistically significant Pearson correlation between age and CD4 count ($R^2 = 0.004$, $P = .18$).

TABLE 3—HIV Risk Factors for Females Whose CD4 Counts Were Tested, South Carolina, April 1989 through September 1990

HIV Risk Factor	Total	No. (%) with CD4 of ≥900
Heterosexual contact		
HIV-positive partner	31	7 (23)
Partner with HIV risk factors	14	4 (29)
Partner without known HIV risk factors	60	10 (17)
Intravenous drug use	13	5 (38)
No reported risk factor	8	3 (38)
Total	126	29

Note. χ^2 association; $P = .31$ for all factors.

Reanalysis of potential risk factors for recent HIV infection with absolute CD4 counts of 800 or 1000 as a cutoff for recent infection or 30 years as a cutoff for age did not substantially change the results.

Discussion

This study of CD4 lymphocyte counts among HIV-infected persons newly diagnosed in public health clinics in South Carolina found that 46% of tested persons had CD4 counts of <500,

the level below which zidovudine may improve disease-free survival.[23–25] This percentage corresponds with (1) the US Army estimate that 58% of their HIV-infected persons have CD4 counts of <500[14] and (2) the 50% of persons with newly diagnosed HIV infections in a Baltimore sexually transmitted disease clinic who had CD4 counts of <500.[13] Although none of these figures is generalizable to the HIV-infected US population, the figures provide evidence across different cohorts that a substantial proportion of HIV-infected persons have far enough advanced HIV disease to potentially benefit from pharmacologic therapy.

Although the majority of persons with HIV infection (even recent HIV infection) were male, a disproportionate number of more recent HIV infections occurred in females. This finding is consistent with other current data. The Centers for Disease Control reports that newly diagnosed cases of AIDS in females increased by 29% from 1988 to 1989 compared with an 18% increase for males.[26] St. Louis and colleagues recently reported US Job Corps data that noted a higher HIV seroprevalence among 16- and 17-year-old females than among males, which the authors suggest may reflect a recent increase in HIV infection in young women nationally.[27]

There are limitations to these results. First, CD4 lymphocyte counts are an imperfect measurement of the duration of HIV infection. There is considerable variability in the normal level of CD4 counts in uninfected persons[17,28] and in the rate of decline of CD4 counts among HIV-infected persons.[2,4] If different groups have different normal values or different rates of decline of CD4 counts, that could distort our extrapolation of the duration of infection from CD4 counts.

However, the limited data that are available suggest that CD4 counts in persons without HIV infection do not differ by gender or ethnicity,[18] and the only data on the effect of age on CD4 counts that we are aware of concern differences in young children, who were excluded from this evaluation. Indeed, if rates of decline of CD4 counts differed among different groups, the duration of infection would be extrapolated most accurately early in the course of the illness before differential rates of decline produced increasingly larger differences (CD4 count decline is linear in individuals with HIV infection[8]). Thus, our data on recently infected persons would be less susceptible to this potential distortion.

A second concern is that the apparent increased prevalence of recent infections in females could reflect their earlier use of health care. More recent infection among females would then reflect earlier diagnosis rather than a trend in disease transmission. However, we have no evidence that females were seeking health care earlier than males. Because the data abstraction methods used did not recover the type of program that a person was tested under, we cannot determine whether recently infected females were disproportionately identified through, for example, prenatal screening efforts, which are not equally accessed by men. We do know, however, that once identified, HIV-infected females were not significantly more likely than males to return within 90 days for follow-up CD4 testing. Continued surveillance of CD4 counts among HIV-infected persons will answer this question. If higher CD4 counts reflect increased rates of new infection among females, then the differences in CD4 counts between the sexes will narrow as the relative numbers of infected females and males stabilize.

A third study limitation is that the study population is not representative of the entire HIV-infected population in South Carolina and so the results may not be generalizable. It included only persons who attended public health clinics and so likely underrepresents higher socioeconomic groups. Further, Blacks and persons who reported no risk factor for HIV transmission were less likely to have follow-up CD4 testing, which could permit an unmeasured selection bias that might affect our findings. Nevertheless, this study population is more diverse than previously reported study populations and so offers the opportunity for a more complete comparison of CD4 counts across age, sex, and HIV risk groups than was previously available.

In conclusion, CD4 counts reflect the duration of HIV infection and so can enhance the characterization of HIV disease in a population. In South Carolina, among HIV-infected persons newly reported to the health department who agreed to CD4 lymphocyte testing, 46% might benefit from immediate drug therapy. Within this study population females may represent an emerging risk group that requires specifically directed HIV prevention efforts. □

Acknowledgments

We are indebted to Robert Ball, MD, MPH, for his work with the CD4 testing program; Pamela Meyer, MSPH, for assistance in data management; Pat Fleming, PhD, Surveillance Section, Division of HIV/AIDS, Centers for Disease Control, for critical review of the manuscript; and Alan Hightower, BA, for assistance in statistical analysis.

References

1. Melbye M, Biggar RJ, Ebbesen P, et al. Long-term seropositivity for human T-lymphotropic virus type III in homosexual men without the acquired immunodeficiency syndrome: development of immunologic and clinical abnormalities. *Ann Intern Med.* 1986;104:496–500.
2. Eyster ME, Gail MH, Ballard JO, Al-Mondhiry H, Goedert JJ. Natural history of immunodeficiency virus infection in haemophilics: effects of T-cell subsets, platelet counts and age. *Ann Intern Med.* 1987;107:1–6.
3. Moss AR, Bacchetti P, Osmond D, et al. Seropositivity for HIV and the development of AIDS or AIDS related condition: three year follow up of the San Francisco General Hospital cohort. *Br Med J.* 1988;296:745–750.
4. Phillips AN, Lee CA, Elford J, et al. Prediction of progression to AIDS by analysis of CD4 lymphocyte counts in a haemophiliac cohort. *AIDS.* 1989;3:737–741.
5. DeWolf F, Lange JA, Houweling JM, et al. Appearance of predictors of disease progression in relationship to the development of AIDS. *AIDS.* 1989;3:563–569.
6. Schwartz K, Visscher B, Detels R, Taylor J, Nishanian P, Fahey J. Immunological changes in lymphadenopathy virus positive and negative symptomless male homosexuals: two years of observation. *Lancet.* 1985;2:831–832.
7. Goedert JJ, Biggar RJ, Melbye M, et al. Effect of T4 count and cofactors on the incidence of AIDS in homosexual men infected with human immunodeficiency virus. *JAMA.* 1987;257:331–334.
8. Phillips AN, Lee CA, Elford J, et al. Serial CD4 lymphocyte counts and development of AIDS. *Lancet.* 1991;337:389–392.
9. Masur H, Ognibene F, Yarchoan R, et al. CD4 counts as predictors of opportunistic pneumonias in human immunodeficiency virus (HIV) infection. *Ann Intern Med.* 1989;111:223–231.
10. Polk BF, Fox R, Brookmeyer R, et al. Predictors of acquired immunodeficiency syndrome developing in a cohort of seropositive homosexual males. *N Engl J Med.* 1987;316:61–66.
11. el-Sadr W, Marmor M, Zolla-Pazner S, et al. Four year prospective study of homosexual males: correlation of immunologic abnormalities, clinical status, and serology to human immunodeficiency virus. *J Infect Dis.* 1987;155:789–793.
12. Eyster ME, Goedert JJ, Sarngadharan

MG, Weiss SH, Gallo RC, Blattner WA. Development and early natural history of HTLV-III antibodies in persons with hemophilia. *JAMA.* 1985;253:2219–2223.
13. Hutchinson CM, Wilson C, Reichart CA, Marsiglia VC, Zenilman JM, Hook EW. CD4 lymphocyte concentrations in patients with newly identified HIV infection attending STD clinics. *JAMA.* 1991;266: 253–256.
14. Brundage JF, McNeil JG, Miller RN, et al. The current distribution of CD4+ T-lymphocyte counts among adults in the United States with human immunodeficiency virus infections: estimates based on the experience of the U.S. Army. *J Acquir Immune Defic Syndr.* 1990;3:92–94.
15. Yanase Y, Tango T, Okumura K, Tada T, Kawasaki T. Lymphocyte subsets identified by monoclonal antibodies in healthy children. *Pediatr Res.* 1986;20:1147–1151.
16. Denny T, Yogev R, Gelman R, et al. Lymphocyte subsets in healthy children during the first 5 years of life. *JAMA.* 1992;267:1484–1488.
17. Malone JL, Simms TE, Gray GC, Wagner KF, Burge JR, Burke DS. Sources of variability in repeated T-helper lymphocyte counts from human immunodeficiency virus type 1-infected patients: total lymphocyte count fluctuations and diurnal cycle are important. *J Acquir Immune Defic Syndr.* 1990;3:144–151.
18. Hill A, Ekong T, Gompels M, Pinching A. CD4 fluctuates within restricted regions of the normal range in heterosexual control subjects. In: *VII International Conference on AIDS Abstract Book.* Rome, Italy: Instituto Superiore Di Sanita; 1991; 1:289.
19. Centers for Disease Control. Table 7. AIDS cases by sex, age at diagnosis, and race/ethnicity, reported through July 1991, United States. *HIV/AIDS Surveill.* August 1991:1–18.
20. Dean AD, Dean JA, Burton JH, Dicker RC. *Epi Info* [computer program]. Version 5.01. Atlanta, Ga: Centers for Disease Control; 1990.
21. *StatXact* [computer program]. Cambridge, Mass: Cytel Software Corporation; 1989.
22. Elliott AC. *Kwikstat* [computer program]. Version 2.00. Cedar Hill, Tex: TexaSoft/ Mission Technologies; 1989.
23. Volberding PA, Lagakos SW, Koch MA, et al. Zidovudine in asymptomatic human immunodeficiency virus infection. *N Engl J Med.* 1990;322:941–949.
24. Graham NMH, Zeger SL, Park LP, et al. The effects on survival of early treatment of human immunodeficiency virus infection. *N Engl J Med.* 1992;326:1037–1042.
25. Aboulker JP, Swart AM. Preliminary analysis of the Concorde trial. *Lancet.* 1993;341:889–890. Letter.
26. Division of HIV/AIDS, Centers for Disease Control. AIDS in females—United States. *MMWR Morb Mortal Wkly Rep.* 1990;39:845–846.
27. St. Louis M, Conway G, Hayman C, Miller C, Petersen L, Dondero T. Human immunodeficiency virus infection in disadvantaged adolescents. *JAMA.* 1991;266: 2387–2391.
28. Weber JN, Wadsworth J, Rogers LA, Moshtael O, Scott K, McManus T. Three year prospective study of HTLV-III/LAV infection in homosexual males. *Lancet.* 1986;i (8491):1179–1182.

The Trading of Sex for Drugs or Money and HIV Seropositivity among Female Intravenous Drug Users

Jacquie Astemborski, MHS, David Vlahov, PhD, Dora Warren, PhD, Liza Solomon, DrPH, and Kenrad E. Nelson, MD

ABSTRACT

Objectives. Data from 538 women in a cohort study recruited in 1988–1989 were analyzed to determine whether trading sex for drugs or money was independently associated with human immunodeficiency virus (HIV) seroprevalence in a population of female intravenous drug users.

Methods. The women were grouped according to the number of partners with whom they reported trading sex for drugs or money during the previous 10 years: none, 1 through 49 (low), or 50 or more (high); the prevalence of HIV seropositivity in the three groups was 23.2%, 23.7%, and 47.6%, respectively. Logistic regression was used to compare the low- and high-trade groups separately with the group that reported no trading.

Results. Low trading was not associated with seroprevalent HIV infection. In a multivariate model, high trading (compared with no trading) was significantly associated with HIV seropositivity after adjustment for cocaine use, history of sexually transmitted diseases, and duration of intravenous drug use.

Conclusions. These data indicate that, among intravenous drug–using women, high levels of trading sex for drugs or money were independently associated with HIV infection. This group needs to be targeted for further intensive intervention. (*Am J Public Health*. 1994; 84:382–387)

Introduction

Intravenous drug users are at increased risk for infection with the human immunodeficiency virus (HIV), owing primarily to the sharing of contaminated injection equipment.[1] Specific risk factors for seroprevalent HIV infection among intravenous drug users include high frequency of injection, use of shooting galleries, and cocaine injection.[2,3] Recent studies have found higher HIV seroincidence rates in female than male intravenous drug users,[4,5] which has led to the suggestion that sexual transmission is more important for women. The extent to which sexual transmission contributes to the occurrence of HIV infection among intravenous drug users has been difficult to demonstrate, with conflicting results across studies.[6–10] A national survey of female prostitutes found significantly higher rates of HIV infection among those who reported intravenous drug use,[11] suggesting that prostitution might be a means of obtaining money to support a drug habit and that the excess HIV infection in these women was probably related to injection behavior. The purpose of this study was to identify whether trading sex for money or drugs is associated with HIV infection among intravenous drug–using women, independent of drug injection practices.

Materials and Methods

Study Population

The organization and data collection methods for this study have been described elsewhere in detail.[12,13] In brief, intravenous drug users were recruited by word of mouth from a variety of community service agencies into a study (the AIDS Link to Intravenous Experiences [ALIVE] Study) of the natural history of HIV infection in a freestanding clinic between February 1988 and March 1989. Eligibility criteria included age older than 17 years, history of injecting drugs within the previous 10 years, and no history of an acquired immunodeficiency syndrome (AIDS)–defining illness at study entry.

Data Collection

After pretest counseling was performed and informed consent was obtained, serum was collected for HIV antibody assays and each participant was interviewed to obtain demographic information; medical history; drug use history, including injection practices, since 1977; and a 10-year history of sexual activity. Participants who reported injecting drugs were asked the last time they had injected. Then all were asked about the following behaviors in the 6 months prior to the last injection: frequency of injection, type of drug used, number of partners shared with, and attendance at shooting galleries. For this study, "current use" was defined as injection within the 6 months prior to interview. We asked all participants about duration of injection in years and then asked for a year-by-year history for the previous 10 years. For women, the 10- year sexual history included the total number of

The authors are with the Department of Epidemiology, The Johns Hopkins School of Public Health, Baltimore, Md.

Requests for reprints should be sent to Jacquie Astemborski, MHS, Room 763, 624 N Broadway, Baltimore, MD 21205.

This paper was accepted September 1, 1993.

different male sex partners, the number of male intravenous drug–using partners, and the number of males with whom sex was traded for drugs or money. In pilot studies, some women were put off by the term "prostitute" but not by the phrase "trading sex for money or drugs"; the latter phrase was adopted for the final instrument used in this study.

Serum samples were assayed for HIV-1 antibodies by enzyme-linked immunosorbent assay (ELISA; Genetic Systems, Seattle, Wash). Confirmation was by Western blot (DuPont Co, Wilmington, Del) techniques.

Data Analysis

Because the risk of HIV infection due to trading sex for drugs or money among women was of primary interest, the sample was restricted to those 538 women for whom this information was available. On the basis of a preliminary analysis, the women were grouped into three categories, according to the number of partners with whom they reported trading sex for drugs or money during the 10 years prior to interview: none, 1 through 49 (low), and 50 or more (high). The cutoff of 50 was chosen to achieve a minimum sample size for meaningful analysis.

For analyses, drug-use variables within the last 6 months were used because recall of specific injection practices was felt to be more accurate for the most recent period and drug injection practices have been shown to be relatively stable over time in this population.[2,3] Therefore, recent behavior combined with duration of injection (more than 5 years vs 5 years or less) were both incorporated into analyses as a means to adjust for drug injection practices when we examined the association of trading sex for drugs or money and HIV seropositivity.

The high and low traders were each compared in separate analyses with the nontraders to examine the association of trading with HIV seroprevalence. Categorical variables were first analyzed by means of odds ratios and chi-square statistics. Variables on univariate analysis that were significant at an alpha level of .05 or less were systematically examined with multiple logistic regression to assess the independent risk associated with each.[14] Variables were retained in the multivariate model if the variables significantly improved the fit of the model.

TABLE 1—Selected Characteristics of Female Intravenous Drug Users (n = 538), by Their Frequency of Trading Sex for Drugs or Money, Baltimore, Md, 1988 through 1989

	No Trade, % (n = 327)	Low Trade, % (1–49 Partners) (n = 169)	High Trade, % (≥ 50 Partners) (n = 42)
HIV seropositive	23.2	23.7	47.6*
Black	91.1	86.4	88.1
Age < 35 y	63.6	74.6*	76.2
> 2 IV drug–using sex partners	10.7	45.6*	78.6*
Any sexually transmitted disease	59.6	78.1*	81.0*
Any oral contraceptive use	58.3	69.7*	71.4
Duration of IV drug use ≥ 5 y	65.8	71.6	83.3*
Current IV drug use[a]	87.8	87.0	81.0
Daily IV drug use[b]	38.5	40.2	45.2
IV cocaine use[c]	76.8	78.1	71.4
Shared needles[d]	57.8	68.1*	53.7
Gallery[e]	12.8	26.0*	47.6*

Note. HIV = human immunodeficiency virus; IV = intravenous.
[a]Used IV drugs within 6 months of interview.
[b]Used IV drugs at least once per day within 6 months of interview.
[c]Used IV cocaine within 6 months of interview.
[d]Shared needles within 6 months of interview.
[e]Visited shooting gallery within 6 months of interview.
*Significantly different from the nontraders at $P \leq .05$.

Results

Of the 552 female intravenous drug users recruited into the ALIVE study, we had complete data on the trading sex question for 538: 327 (60.8%) reported no trading, 169 (31.4%) traded with fewer than 50 partners, and 42 (7.8%) traded with 50 or more partners. The median number of male sex partners (including trading partners) in the previous 10 years for the no-, low-, and high-trade groups was 3, 8, and 400, respectively.

The prevalence of HIV seropositivity among the nontraders, low traders, and high traders was 23.2%, 23.7% and 47.6%, respectively (Table 1). Compared with the nontraders, the low traders were younger and were more likely to have had sex with more than two intravenous drug users, to have had a sexually transmitted disease, and to have used oral contraceptives. Low traders were more likely than intravenous drug users who did not trade to have shared needles and to have gone to shooting galleries within the 6 months prior to interview, but they were not significantly different from those who had never traded with respect to current injection drug use, frequency of injection, or intravenous use of cocaine.

High traders were more likely than nontraders to have had sex with more than two intravenous drug users, to have had a sexually transmitted disease, and to have used intravenous drugs for at least 5 years. High traders did not differ from nontraders with regard to frequency of current intravenous drug use, frequency of injection, intravenous cocaine use, or having shared needles, but they were more likely to have attended shooting galleries.

When the analysis was restricted to the low- and no-trade groups (Table 2), univariate analysis showed that having traded sex for drugs or money was not associated with HIV seropositivity. Factors associated with HIV seropositivity were being Black, having had sex with more than two intravenous drug users, history of drug treatment, having used intravenous drugs for 5 or more years, having used intravenous cocaine, and having shared needles with more than one partner. As Table 3 shows, factors independently associated with HIV seropositivity on multivariate analysis were being Black (odds ratio [OR] = 4.75), having had sex with more than two intravenous drug users (OR = 2.31), and having used intravenous cocaine (OR = 3.10).

When the analysis was restricted to the high- and no-trade groups (Table 4), univariate analysis showed having traded sex for drugs or money to be significantly associated with HIV seropositivity (OR = 3.00). Factors associated with

TABLE 2—Unadjusted Odds Ratios for Seroprevalent HIV among Female Intravenous Drug Users Who Reported No or Low Trading of Sex for Money or Drugs (n = 496), Baltimore, Md, 1988 through 1989

	n	% Seropositive for HIV	OR	P
Frequency of trading sex for money or drugs				
Low (1–49 partners)	169	23.7	1.02	
None	327	23.2	1.00	.915
Race				
Black	444	25.5	5.58	
Other	52	5.8	1.00	.002
No. IV drug–using sex partners[a]				
>2	112	34.8	2.12	
0–2	383	20.1	1.00	.001
Drug treatment				
Ever	289	26.6	1.57	
Never	207	18.8	1.00	.043
Duration of IV drug use				
≥5 y	336	26.8	1.89	
<5 y	160	16.3	1.00	.010
Current IV drug use[b]				
Yes	434	24.7	1.93	
No	62	14.5	1.00	.078
IV cocaine use[c]				
Yes	383	27.4	3.50	
No	113	9.7	1.00	<.001
Shared with >1 partner[d]				
Yes	177	29.4	1.66	.019
No	319	20.1	1.00	

Note. HIV = human immunodeficiency virus; IV = intravenous; OR = odds ratio.
[a]Data are missing for 1 subject.
[b]Used IV drugs within 6 months of interview.
[c]Used IV cocaine within 6 months of interview.
[d]Shared needles within 6 months of interview.

TABLE 3—Adjusted Odds Ratios for Seroprevalent HIV in a Population of Female Intravenous Drug Users Who Reported No or Low Trading of Sex for Money or Drugs (n = 496), Baltimore, Md, 1988 through 1989

	OR	95% CI
Race		
Black	4.75	1.42, 15.93
Other	1.00	
No. IV drug–using sex partners		
>2	2.31	1.43, 3.74
0–2	1.00	
Current IV cocaine use[a]		
Yes	3.10	1.57, 6.09
No	1.00	

Note. HIV = human immunodeficiency virus; IV = intravenous; OR = odds ratio; CI = confidence interval. Low trading was defined as trading with from 1 through 49 partners.
[a]Used IV cocaine within 6 months of interview.

being HIV positive on univariate analysis were being Black, age less than 35 years, having had sex with more than two intravenous drug users, history of any sexually transmitted disease, having used drugs for at least 5 years, having injected drugs within the last 6 months, having injected drugs at least once per day, having used intravenous cocaine, and having shared needles with more than one partner. Lesbian and bisexual sex practices were examined and found not to be associated with HIV seropositivity (OR = 1.13). As Table 5 shows, factors independently associated with HIV seropositivity on multivariate analysis were having traded sex for drugs or money (OR = 2.99), history of any sexually transmitted disease (OR = 1.84), 5 or more years of intravenous drug use (OR = 1.87), and having used intravenous cocaine (OR = 3.54). Additional analyses using continuous rather than categorical exposure variables did not alter the primary association of interest (data not shown).

In the analysis that was restricted to the high- and low-trade groups, a high frequency of having traded sex for drugs or money was associated with HIV seropositivity (OR = 2.93). Other factors associated with being HIV seropositive on univariate analysis were having had sex with more than two intravenous drug users, having undergone drug treatment, having used intravenous drugs for at least 5 years, having been arrested in the last 10 years, and having used intravenous cocaine. On multivariate analysis, only having traded sex for drugs or money (high trade) and having had sex with more than two intravenous drug users were associated with HIV seropositivity. Additionally, there was an interaction between having traded sex and having had sex with two or more intravenous drug users. Within the low-trade subset, women who had had sex with more than two intravenous drug users had an odds ratio for HIV seropositivity of 3.74 (95% confidence interval [CI] = 1.74, 8.05) compared with those who had had sex with two or fewer intravenous drug users. Within the high-trade subset, women who had had sex with more than two intravenous drug users had an odds ratio for HIV seropositivity of 0.37 (95% CI = 0.08, 1.73) compared with those who had had sex with two or fewer intravenous drug users. When only women who reported having had sex with two or fewer intravenous drug users were considered, high traders had an odds ratio for HIV seropositivity of 12.83 (95% CI = 2.83, 58.30).

Discussion

The data presented here are consistent with earlier published reports noting that the exchange of sex for money or drugs is common among intravenous drug–using women.[15–22] In this study, nearly 40% of the women reported exchanging sex for drugs or money at least once in the previous 10 years.

TABLE 4—Unadjusted Odds Ratios for Seroprevalent HIV in a Population of Female Intravenous Drug Users Who Reported No or High Trading of Sex for Money or Drugs (n = 369), Baltimore, Md, 1988 through 1989

	n	% Seropositive for HIV	OR	P
Frequency of trading sex for money or drugs				
High (≥ 50 partners)	42	47.6	3.00	
None	327	23.2	1.00	.001
Race				
Black [a]	335	27.5	2.84	
Other	34	11.8	1.00	.047
Age				
< 35 y	240	30.0	1.88	
> 35 y	129	18.6	1.00	.018
No. IV drug–using sex partners[a]				
> 2	68	36.8	1.88	
0–2	300	23.7	1.00	.026
Any sexually transmitted disease				
Yes	229	31.4	2.22	
No	140	17.1	1.00	.002
Duration of IV drug use				
≥ 5 y	250	30.4	2.16	
< 5 y	119	16.8	1.00	.005
Current IV drug use[b]				
Yes	321	28.0	2.73	
No	48	12.5	1.00	.022
Current use ≥ 1 day[c]				
Yes	145	33.1	1.81	.013
No	224	21.4	1.00	
Current IV cocaine use[d]				
Yes	281	30.6	3.44	
No	88	11.4	1.00	< .001
Shared with > 1 partner[a,e]				
Yes	111	36.0	2.02	
No	257	21.8	1.00	.004

Note. HIV = human immunodeficiency virus; IV = intravenous; OR = odds ratio.
[a]Data are missing for 1 subject.
[b]Used IV drugs within 6 months of interview.
[c]Used IV drugs at least once per day within 6 months of interview.
[d]Used IV cocaine within 6 months of interview.
[e]Shared needles within 6 months of interview.

TABLE 5—Adjusted Odds Ratios for Seroprevalent HIV in a Population of Female Intravenous Drug Users Who Reported No or High Trading of Sex for Money or Drugs (n = 369), Baltimore, Md, 1988 through 1989

	OR	95% CI
Frequency of trading sex for money or drugs		
High (≥ 50 partners)	2.99	1.47, 6.10
None	1.00	
Any sexually transmitted disease		
Yes	1.84	1.07, 3.16
No	1.00	
Duration of IV drug use		
≥ 5 y	1.87	1.05, 3.16
< 5 y	1.00	
Current IV cocaine use[a]		
Yes	3.54	1.69, 7.40
No	1.00	

Note. HIV = human immunodeficiency virus; IV = intravenous; OR = odds ratio; CI = confidence interval.
[a]Used IV cocaine within 6 months of interview.

A major finding was a significantly higher rate of HIV infection for high traders (≥ 50 partners) than for nontraders. This statistical association persisted after simultaneous adjustment for multiple confounders, including duration of drug injection, use of intravenous cocaine, and history of sexually transmitted diseases. A significant association of HIV infection and trading sex for money or drugs after adjustment for drug injection practices indicates that the relationship was probably not due to more frequent or riskier drug injection practices among the high traders. The fact that a history of sexually transmitted diseases was independently associated with HIV infection after drug injection variables were controlled is consistent with a role for sexual transmission of HIV infection in this population. The statistically significant association of HIV seropositivity and high trading after adjustment for history of sexually transmitted diseases indicates that there is an independent effect of frequent trading of sex for money or drugs— that the behavior of trading is risky irrespective of history of sexually transmitted disease.

An unexpected finding was that HIV seropositivity rates were similar between the low traders (1–49 partners) and the nontraders. Because the median number of sex trading partners in the low-trade group was five, and the total median number of sex partners (including trading partners) in the previous 10 years was similar between the two groups (three for the no- trade group and eight for the low-trade group), the groups might be considered relatively homogeneous with respect to number of sex partners. It is also conceivable that the question about trading sex for money or drugs might have been perceived differently by different study participants; for example, women who reported a low number of trading partners might have included instances where they injected with a partner immediately prior to sex, but other women might not have considered such instances, especially with a regular sex partner, to be an exchange of sex for drugs. Additional studies will need to form questions differently to explore semantic and actual variations.

In the analysis that was restricted to low traders and nontraders, HIV infec-

tion was significantly associated with being Black, having used intravenous cocaine, and having had sex with more than two male intravenous drug users. Although similar findings have been reported by others for injection drug users as a group,[2,3,8,23] the difference here is that these variables distinguished among subgroups of intravenous drug–using women who had similar HIV seroprevalence rates. The data suggest differing risk profiles within a single population drawn from the same city during the same calendar time. Higher rates of HIV seropositivity among racial/ethnic minorities have been observed in other studies and probably reflect differences in risk behaviors that are inadequately measured with current instruments.[24]

HIV seroprevalence was higher in the high-trade group than in the low-trade group. In the analysis of risk factors involving these two groups, a significant interaction term showed that having had sex with two or more male intravenous drug users was risky for the low-trade group but, paradoxically, was "protective" for the high-trade group. This finding is consistent with the interpretation that women in the high-trade group were probably less likely to know whether their sex partners were high-risk and therefore less likely to report them. In addition, high traders and low traders may associate with different networks of individuals within which the underlying seroprevalence rates differ. More detailed data, including type of prostitution and frequency of condom use, are needed to better disentangle the effects of drug injection and sexual transmission within different social networks.

Before firm conclusions are drawn, several study limitations should be acknowledged. Because the sample was composed of volunteers, the extent to which this study population is representative of intravenous drug users in Baltimore or elsewhere is unclear. Therefore, this relationship should be examined in other cities with varying seroprevalence rates. The cross-sectional design, although efficient, cannot distinguish risk factors from prevalence correlates; therefore, prevalence-incidence bias cannot be excluded. Additional prospective studies are needed to confirm and expand the findings reported here.

In conclusion, although early studies of intravenous drug users found that HIV seroprevalence was similar for men and women[2,25] and some investigators concluded that transmission of HIV infection among intravenous drug users was parenteral rather than sexual,[6,7] more recent studies have acknowledged a more prominent role of sexual transmission in this population.[8,9] Further support of the importance of sexual transmission is provided by studies of incident HIV infection among current and former intravenous drug users that have shown higher rates of HIV infection in women than in men.[4,5] Other investigators have found an association between crack use and HIV seroprevalence[26,27] that is thought to be mediated by increased sexual activity. This study contributes to the growing literature on risk factors for HIV infection among female intravenous drug users by identifying an independent risk associated with the practice of trading sex for money or drugs in this population. Multiple studies have found that although intravenous drug users have changed drug-using behaviors in response to the AIDS epidemic, progress in the prevention of sexual transmission among intravenous drug users has not kept pace.[28] The data presented here provide additional evidence to support redoubled efforts to intervene on the issue of sexual transmission of HIV infection in this vulnerable population. □

Acknowledgments

This research was supported by grants DA 04334 and DA 05911 from the National Institute on Drug Abuse.

We are grateful to the staff and participants of the ALIVE study, without whom this work would not have been possible, and to Dr Anneke van den Hoek, who reviewed an earlier version of the manuscript.

References

1. Friedland GH, Klein RS. Transmission of the human immunodeficiency virus. *N Engl J Med.* 1987;31:1125–1135.
2. Vlahov D, Muñoz A, Anthony JC, Cohn S, Celentano DD, Nelson KE. Association of drug injection patterns with antibody to human immunodeficiency virus type 1 among intravenous drug users in Baltimore, Maryland. *Am J Epidemiol.* 1990;132:847–856.
3. Anthony JC, Vlahov D, Nelson KE, Cohn S, Astemborski J, Solomon L. New evidence on intravenous cocaine use and the risk of infection with the human immunodeficiency virus type 1. *Am J Epidemiol.* 1991;134:1175–1189.
4. Nelson KE, Vlahov D, Muñoz A, Cohn S, Solomon L. Incident HIV-1 infection in a cohort of intravenous drug users. Presented at the Eighth International Conference on AIDS/Third STD World Congress; July 19–24, 1992; Amsterdam, the Netherlands. Abstract No. POC 4698.
5. Nicolosi A, Correa Leite ML, Musicco M, Molinari S, Lazzarin A, for the Northern Italian Seronegative Drug Addicts (NISDA) Study. Parenteral and sexual transmission of human immunodeficiency virus in intravenous drug users: a study of seroconversion. *Am J Epidemiol.* 1992; 135:225–233.
6. Battjes RJ, Pickens RW, Amsel Z, Brown LS Jr. Heterosexual transmission of human immunodeficiency virus among intravenous drug users. *J Infect Dis.* 1990;162: 1007–1011.
7. Tirelli U, Vacchev E, Carbone A, De-Paeli P, Monfardini S. Heterosexual contact is not the predominant mode of HTLV-III transmission among drug users. *JAMA.* 1986;255:2289. Letter.
8. Schoenbaum EE, Hartel D, Selwyn PA, et al. Risk factors for human immunodeficiency virus infection in intravenous drug users. *N Engl J Med.* 1989;321:874–879.
9. Salmaso S, Conti S, Sasse H, and the Second Multicenter Study Group on Drug Users. Drug use and HIV-1 infection: report from the second Italian multicenter study. *J Acquir Immune Defic Syndr.* 1991;4:607–613.
10. Nelson KE, Vlahov D, Cohn S, et al. Sexually transmitted diseases in a population of intravenous drug users: association with seropositivity to the human immunodeficiency virus (HIV). *J Infect Dis.* 1991;164: 457–463.
11. Centers for Disease Control. Antibody to HIV in female prostitutes. *MMWR.* 1987; 36:157–161.
12. Vlahov D, Anthony JC, Muñoz A, et al. The ALIVE Study, a longitudinal study of HIV-1 infection in intravenous drug users: description of methods and characteristics of participants. In: Hartsock PI, Genser SG, eds. *Longitudinal Studies of HIV Infection in intravenous drug users: methodological issues in natural history research.* Rockville, Md: US Public Health Service; 1991:75–100. DHHS publication ADM 91-1786.
13. Anthony JC, Vlahov D, Celentano DD, et al. Self-report interview data for a study of HIV-1 infection among intravenous drug users: description of methods and preliminary evidence on validity. *J Drug Issues.* 1991;21:739–757.
14. Breslow NE, Day NE. *Statistical Methods in Cancer Research, Vol 1: The Analysis of Case–Control Studies.* Lyons, France: International Agency for Research on Cancer; 1980.
15. Franceschi S, Tirelli U, Vaccher E, et al. Risk factors for HIV infection in drug addicts from the northeast of Italy. *Int J Epidemiol.* 1988;17:162–167.
16. Lewis DK, Watters JK. Sexual risk behavior among heterosexual intravenous drug users: ethnic and gender variations. *AIDS.* 1991;5:77–83.
17. Saxon AJ, Calsyn DA, Whittaker S, Freeman G Jr. Sexual behaviors of intravenous drug users in treatment. *J Acquir Immune Defic Syndr.* 1991;4:938–944.

18. Booth R, Koester S, Brewster JT, Weibel WW, Fritz RB. Intravenous drug users and AIDS: risk behaviors. *Am J Drug Alcohol Abuse.* 1991;17:337–353.
19. Williams ML. A model of the sexual relations of young I.V. drug users. *J Acquir Immune Defic Syndr.* 1990;3:192. Letter.
20. Chaisson RE, Moss AR, Onishi R, Osmond D, Carlson JR. Human immunodeficiency virus infection in heterosexual intravenous drug users in San Francisco. *Am J Public Health.* 1987;77:169–172.
21. van den Hoek JAR, Coutinho RA, van Haastrecht HJA, Zadelhoff AW, Goudsmit J. Prevalence and risk factors of HIV infection among drug users and drug-using prostitutes in Amsterdam. *AIDS.* 1988;2: 55–60.
22. van den Hoek JAR, van Haastrecht HJA, Scheeringa-Troost B, Goudsmit J, Coutinho RA. HIV infection and STD in drug addicted prostitutes in Amsterdam: potential for heterosexual HIV transmission. *Genitourin Med.* 1989;65:146–150.
23. Chaisson RE, Bacchetti P, Osmond D, Brodie B, Sande MA, Moss AR. Cocaine use and HIV infection in intravenous drug users in San Francisco. *JAMA.* 1989;261: 561–565.
24. Samuels J, Vlahov D, Anthony JC, Chaisson RE. Measurement of risk behaviors among intravenous drug users. *Br J Addict.* 1992;87:417–428.
25. Marmor M, Des Jarlais DC, Cohen H, et al. Risk factors for infection with human immunodeficiency virus among intravenous drug abusers in New York City. *AIDS.* 1987;1:39–44.
26. Chiasson MA, Stoneburner RL, Hildebrandt DS, Ewing WE, Telzak EE, Jaffe HW. Heterosexual transmission of HIV-1 associated with the use of smokable freebase cocaine (crack). *AIDS.* 1991;5: 1121– 1126.
27. Chirgwin K, DeHovitz JA, Dillon S, McCormack WM. HIV infection, genital ulcer disease, and crack cocaine use among patients attending a clinic for sexually transmitted diseases. *Am J Public Health.* 1991; 81:1576–1579.
28. Centers for Disease Control. Risk behaviors for HIV transmission among intravenous drug users not in drug treatment—United States 1987–1989. *MMWR.* 1990; 39:273–276.

Changes in HIV-Related Information Sources, Instruction, Knowledge, and Behaviors among US High School Students, 1989 and 1990

ABSTRACT

Objectives. Few data have been available among adolescents to determine behavioral changes that may prevent human immunodeficiency virus (HIV) infection. This analysis examines changes in the prevalence of self-reported HIV-related information sources, instruction, knowledge, and behaviors among high school students in the United States.

Methods. Two independent, multistage national probability samples of students in grades 9 through 12 were surveyed in 1989 (n = 8098) and 1990 (n = 11 631) with self-administered, anonymous questionnaires that included similar items.

Results. Compared with students surveyed in 1989, a significantly greater proportion of students surveyed in 1990 had received HIV instruction in school. Significant decreases were found in the proportion of White and female students who reported having had sexual intercourse, in the proportion of White students reporting two or more lifetime sex partners, and in the proportion of 15- and 16-year-olds, White students, and female students who reported having had four or more lifetime sex partners. For both years, students who had a greater level of HIV knowledge were less likely to have had multiple lifetime sex partners or to have injected illicit drugs.

Conclusions. The findings suggest that school-based HIV education and knowledge may be contributing factors in reducing certain risk behaviors that can lead to HIV transmission among secondary school youth. (*Am J Public Health.* 1994;84:388–393)

Deborah Holtzman, PhD, Richard Lowry, MD, MS, Laura Kann, PhD, Janet L. Collins, PhD, and Lloyd J. Kolbe, PhD

Introduction

Since the latter half of the 1980s, schools throughout the United States have been offering education programs to teach adolescents about human immunodeficiency virus (HIV) infection. However, the effect of such programs on HIV-related knowledge and behaviors among adolescents is largely unknown. Several recent studies have investigated the extent to which adolescents in the United States engage in behaviors that place them at risk for HIV infection.[1–5] Although a few studies have tried to assess changes in these behaviors over time, their findings have been limited.[6–8]

To begin to evaluate the effect of national public health information and education efforts on HIV-related behaviors of US adolescents, we compared data from two national probability samples of students in grades 9 through 12. In 1989, the Centers for Disease Control (CDC) conducted the Secondary School Student Health Risk Survey[4] to measure the prevalence of HIV-related knowledge, beliefs, and behaviors among US high school students. In 1990, using a similar methodology, CDC's Youth Risk Behavior Survey expanded this effort by measuring priority health-risk behaviors for HIV infection and other major causes of morbidity and mortality.[9] Eight HIV-related questions were either identical on the two surveys or were similar enough to allow comparisons.

In this analysis we examine the prevalence of HIV-related sources of information, instruction, and knowledge and the prevalence of selected HIV-related sexual and drug use behaviors and assess changes in these rates from 1989 to 1990. We also examine relationships among the variables in 1990 to determine predictors of HIV-related sexual and drug use behavior and compare these findings with those from analyses of the 1989 data.[3,5] We hypothesized that greater levels of HIV knowledge would be negatively related to the risk behaviors and that the relationships would not change between the two time periods.

Methods

Study Design

By use of self-administered questionnaires, data were obtained from two independent samples of high school students who were surveyed in spring 1989 and spring 1990, respectively. Both were designed as probability samples of all 9th through 12th grade public and private school students in the United States, Puerto Rico, and the Virgin Islands. Schools were randomly selected within primary sampling units (counties or groups of counties), one or two classes per grade were randomly selected within schools, and all students within selected classes were eligible to participate. The sampling frames were stratified by geographic region and metropolitan status and additionally by race/ethnicity in the

The authors are with the Division of Adolescent and School Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Ga.

Requests for reprints should be sent to Deborah Holtzman, PhD, Division of Adolescent and School Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Mailstop K-33, 4770 Buford Hwy, NE, Atlanta, GA 30341-3724.

This paper was accepted June 2, 1993.

1990 Youth Risk Behavior Survey. Both samples were selected with probability proportional to enrollment size. Schools with substantial numbers of Black and Hispanic students were oversampled to ensure adequate numbers of students in these two groups. Weighting factors were developed to compensate for unequal selection probabilities and to adjust for nonresponse in both surveys.

Additional information about the 1989 Secondary School Student Health Risk Survey sample can be found elsewhere.[3-5] Like the 1989 survey, the 1990 Youth Risk Behavior Survey questionnaire was designed for a seventh-grade reading level and administered in the classroom by professionally trained data collectors. Parents provided consent for students to participate, and student anonymity was protected.

Data Analysis

We used *t* tests to determine significant differences between 1989 and 1990 for selected student responses. Following previous studies of the 1989 Secondary School Student Health Risk Survey,[3,5] we examined the 1990 Youth Risk Behavior Survey data to determine relationships between two dependent variables (sexual behavior [having had two or more sex partners during a lifetime] and injecting drug use [having ever injected illicit drugs]) and two independent variables (having received HIV instruction [yes/no] and having correct knowledge about HIV [based on responses to two items]). (In establishing models for the 1990 survey data, we made every effort to replicate the two previous analyses of the 1989 survey data. However, in the 1989 analyses, 17 items—instead of 2—were available to measure HIV knowledge. Also the models for the 1989 survey included the interactions between HIV instruction and sex and between instruction and race/ethnicity. We excluded these interaction terms from our models for the 1990 survey because of the relatively high correlations between the terms and their main effects.) We used stepwise multiple logistic regression and controlled for age, sex, and race/ethnicity in each model.

All analyses were conducted on weighted data. We calculated confidence intervals (CIs) and statistical tests by using SESUDAAN procedures,[10] which account for the complex sample design. Results were considered significant if the *P* value for the calculated statistic was less than or equal to .05.

TABLE 1—Percentage of Students Who Received HIV Instruction in School, by Demographic Characteristic

	1989 Secondary School Student Health Risk Survey			1990 Youth Risk Behavior Survey		
	%	95% CI	No. Observations	%	95% CI	No. Observations
Age, y						
<15	57.7	49.6, 65.8	697	74.8	68.1, 81.6*	1534
15–16	57.0	50.9, 63.1	3957	75.2	70.2, 80.2*	5507
>16	49.1	42.3, 56.0	3432	72.4	66.9, 77.9*	4522
Race/ethnicity						
White	54.0	47.5, 60.5	5102	73.8	67.8, 79.7*	6302
Black	54.9	47.2, 62.7	1153	79.2	75.5, 82.9*	2262
Hispanic	51.4	43.1, 59.6	1366	76.1	72.1, 80.2*	2253
Other[a]	51.4	41.5, 61.2	460	61.9	53.4, 70.3	722
Sex						
Female	54.0	48.0, 60.1	4335	75.6	70.9, 80.2*	5928
Male	53.4	47.6, 59.2	3745	72.5	67.2, 77.9*	5632
Total	53.7	48.0, 59.5	8088	74.0	69.1, 79.0*	11 567

Source. The percentages and confidence intervals for the 1989 survey are reproduced with permission from Holtzman et al.[5]

[a]Includes Native Americans, Alaskan Natives, Asian or Pacific Islanders, and persons of other unspecified groups.

*$P \leq .05$.

TABLE 2—Percentage of Students Who Reported Having Had Sexual Intercourse, by Demographic Characteristic

	1989 Secondary School Student Health Risk Survey		1990 Youth Risk Behavior Survey	
	%	95% CI	%	95% CI
Age, y				
<15	33.5	27.4, 39.5	35.0	30.2, 39.9
15–16	52.9	48.9, 56.9	47.7	43.5, 51.9
>16	69.7	66.2, 73.2	68.0	65.0, 70.9
Race/ethnicity				
White	57.9	54.9, 60.9	51.6	48.7, 54.5*
Black	77.1	72.5, 81.6	72.3	68.5, 76.1
Hispanic	51.8	46.9, 56.7	53.4	48.6, 58.1
Other[a]	43.3	35.4, 51.1	43.7	33.9, 53.5
Sex				
Female	53.8	50.5, 57.0	48.0	45.3, 50.7*
Male	63.1	59.5, 66.8	60.8	56.5, 65.1
Total	58.5	55.5, 61.5	54.2	51.2, 57.3

Source. The race and sex percentages and confidence intervals for the 1989 survey are reproduced with permission from Kann et al.[4,p.3369]

[a]Includes Native Americans, Alaskan Natives, Asian or Pacific Islanders, and persons of other unspecified groups.

*$P \leq .05$.

Results

Of 122 schools selected for the 1989 Secondary School Student Health Risk Survey, 99 (81%) participated. Usable questionnaires were received from 8098 (83%) of the eligible students. Of 168 schools selected for the 1990 Youth Risk Behavior Survey, 124 (74%) participated, and 11 631 (87%) students pro-

TABLE 3—Percentage of Students Who Reported Having Had Sexual Intercourse with Two or More Partners during Their Lifetime, by Demographic Characteristic

	1989 Secondary School Student Health Risk Survey		1990 Youth Risk Behavior Survey	
	%	95% CI	%	95% CI
Age, y				
<15	21.7	16.1, 27.3	23.0	19.2, 26.7
15–16	34.6	30.3, 38.8	30.8	27.5, 34.2
>16	50.1	45.9, 54.3	46.9	43.5, 50.4
Race/ethnicity				
White	38.5	35.2, 41.8	32.4	29.7, 35.1*
Black	63.9	58.3, 69.4	59.8	55.8, 63.8
Hispanic	36.0	31.6, 40.4	34.6	29.7, 39.5
Other[a]	28.5	20.8, 36.2	29.2	21.0, 37.5
Sex				
Female	33.4	30.1, 36.7	29.4	26.7, 32.0
Male	46.7	42.7, 50.8	43.6	39.6, 47.5
Total	40.1	36.8, 43.5	36.3	33.4, 39.2

Sources. The race and sex percentages and confidence intervals for the 1989 survey are reproduced from Anderson *et al.*[7] with the permission of the Alan Guttmacher Institute.

[a]Includes Native Americans, Alaskan Natives, Asian or Pacific Islanders, and persons of other unspecified groups.

*$P \leq .05$.

TABLE 4—Percentage of Students Who Reported Having Had Sexual Intercourse with Four or More Partners during Their Lifetime, by Demographic Characteristic

	1989 Secondary School Student Health Risk Survey		1990 Youth Risk Behavior Survey	
	%	95% CI	%	95% CI
Age, y				
<15	12.7	8.6, 16.8	12.0	9.1, 14.8
15–16	20.0	17.1, 23.0	14.7	12.6, 16.8*
>16	29.9	26.5, 33.2	26.4	23.4, 29.5
Race/ethnicity				
White	22.0	19.9, 24.2	15.8	13.8, 17.8*
Black	43.6	38.4, 48.8	37.8	33.0, 42.6
Hispanic	20.4	15.7, 25.0	16.5	13.4, 19.6
Other[a]	16.9	10.7, 23.2	16.8	11.3, 22.3
Sex				
Female	16.0	13.5, 18.6	11.8	10.3, 13.3*
Male	31.1	27.9, 34.3	26.7	23.0, 30.4
Total	23.6	21.2, 26.0	19.1	17.2, 21.0*

[a]Includes Native Americans, Alaskan Natives, Asian or Pacific Islanders, and persons of other unspecified groups.

*$P \leq .05$.

vided usable data. Comparisons of selected and participating schools by region, metropolitan status, and race/ethnicity (in the 1989 survey)[5] and by region and metropolitan status (in the 1990 survey) showed small differences within each category, suggesting minimal nonresponse bias among the samples.

Demographic characteristics of the samples for both years showed that participants were distributed almost evenly by grade and by sex. With two exceptions, the samples also had similar distributions by age group and by race/ethnicity. The 1989 and 1990 samples differed in the weighted proportion of Black students (9.1% vs 15.2%) and the weighted proportion of students younger than 15 years of age (8.4% vs 13.1%). To determine whether these differences would influence results of the *t*-test analysis, we conducted additional multivariate analyses that took into account the demographic characteristics of the samples (data not shown). Controlling for age, sex, and race/ethnicity, we regressed each HIV-related measure on time. All significant changes observed in the *t*-test analysis were confirmed by the multivariate analysis, indicating that the small differences in the samples did not influence the changes we found in reported instruction, knowledge, or behavior from 1989 to 1990. (In fact, differences from 1989 to 1990 that were not statistically significant in the *t*-test analysis were significant in the multivariate analysis. Thus, the findings reported here from the *t*-test analysis are statistically more conservative.) Furthermore, the weighted percentages of White, Black, and Hispanic students in both surveys were not unlike the racial and ethnic distribution of 9th through 12th grade students in the United States, suggesting that the samples were representative of the general population of students.[5]

Instruction, Knowledge, and Behavior

In both years, students were asked whether they had been taught about the acquired immunodeficiency syndrome (AIDS) and HIV infection in school. The percentage of students receiving HIV instruction significantly increased from 53.7% in 1989 to 74.0% in 1990; these increases were evident for almost every demographic subgroup (Table 1). Moreover, a significantly greater proportion of students in 1990 (68.4%; 95% CI = 65.0, 71.7) than in 1989 (49.0%; 95% CI = 46.5, 51.5) reported that they knew where to get good information about AIDS/HIV infection. However, students were no more likely to talk about AIDS/HIV infection with their parents or other adult family members in 1990 (52.8%; 95% CI = 50.9, 54.8) than they were in 1989 (54.4%; 95% CI = 51.9, 56.8). And although students in 1990 (61.1%; 95% CI = 56.5, 65.6) were significantly more likely than those in 1989 (45.1%; 95% CI = 42.3, 48.0) to know that insects do not transmit the virus, significantly fewer students in 1990 (54.9%; 95% CI = 52.1, 57.7) than in 1989 (64.0%; 95% CI = 61.3, 66.7) reported that they knew the virus could not be transmitted from donating blood.

Essentially no change was observed from 1989 (8.1%; 95% CI = 6.6, 9.6) to 1990 (8.3%; 95% CI = 6.9, 9.7) in the proportion of students who reported that they were younger than age 13 when they first had sexual intercourse. However, in 1990, proportionately fewer students in most subgroups reported that they had had sexual intercourse (Table 2). Notably, these decreases were significant among White students and female students.

Similarly, the proportion of all students who reported having had two or more sex partners during their lifetime was lower in 1990 (36.3%) than in 1989 (40.1%) (Table 3). Although this decline was evident by age group, racial/ethnic group, and sex, with one exception the differences were not statistically significant. Among White students, a significantly smaller proportion reported two or more lifetime sex partners in 1990 than in 1989. For reported sexual intercourse with four or more partners, however, the decrease in prevalence between 1989 and 1990 was significant for all students; within subgroups, the decline was significant for 15- and 16-year-olds, White students, and female students (Table 4).

In 1990, 1.5% (95% CI = 1.0, 1.9) of students reported having ever injected illicit drugs. We were unable, however, to directly compare this estimate with that for 1989. Although the wording of the relevant questions was similar in 1989 and 1990, different criteria were used to define injecting drug use in the 1990 survey.[11] Likewise, we were unable to compare condom use from 1989 to 1990; however, for this behavior, the wording of the relevant questions differed. In 1989, students were asked about the frequency of condom use whenever they had sexual intercourse. In 1990, students were asked whether they had used a condom at last sexual intercourse.

Results of the multivariate analysis of the 1990 Youth Risk Behavior Survey data revealed that having lower levels of HIV knowledge, being male, being Black, and being in an older age group were all significantly associated with having had two or more lifetime sex partners (Table 5). A previous study of the 1989 Secondary School Student Health Risk Survey found that these same variables were significantly associated with having had two or more lifetime sex partners.[3] Importantly, after we controlled for demographic characteristics, students with lower levels of HIV knowledge were more likely to report two or more lifetime sex partners.

TABLE 5—Logistic Regression Analysis of Factors Predicting Multiple Sex Partners, 1990 Youth Risk Behavior Survey

Factor[a]	*b*	Standard Error (*b*)	Odds Ratio	95% CI
HIV knowledge	−0.111*	0.045	0.90	0.82, 0.98
Male	0.632*	0.074	1.88	1.63, 2.18
Age	0.372*	0.029	1.45	1.37, 1.54
Black	1.338*	0.097	3.81	3.15, 4.61
Hispanic	0.182	0.131	1.20	0.93, 1.55

[a]Reference categories for male and for Black and Hispanic are female and other racial/ethnic group, respectively.
*$P < .05$.

TABLE 6—Logistic Regression Analysis of Factors Predicting Ever Injecting Drugs, 1990 Youth Risk Behavior Survey

Factor[a]	*b*	Standard Error (*b*)	Odds Ratio	95% CI
HIV instruction	−0.356	0.293	0.70	0.39, 1.24
HIV knowledge	−0.274*	0.129	0.76	0.59, 0.98
Male	1.221*	0.268	3.39	2.01, 5.73
White	−0.530	0.269	0.59	0.35, 1.00
Black	−1.549*	0.439	0.21	0.09, 0.50

[a]Reference categories for HIV instruction, male, and White and Black are no HIV instruction, female, and other racial/ethnic group.
*$P < .05$.

Our analysis of the 1990 survey data also found that having lower levels of HIV knowledge, being male, and not being Black compared with other race/ethnic groups were all associated with reported lifetime drug injection (Table 6). Again, these results were consistent with those reported for 1989.[5] Notably, controlling for demographic characteristics, we found that students in 1990 who reported lower levels of HIV knowledge also were more likely to report ever injecting drugs. Lastly, although HIV instruction did not directly affect these behavioral outcomes, when we controlled for age, sex, and race/ethnicity, HIV instruction was positively associated with HIV knowledge in analyses of the 1990 data (similar to our findings for the 1989 data[5]).

Discussion

Programs to educate students about HIV have been implemented in many of the nation's schools.[12] But whether such programs are associated with improvements in HIV knowledge and changes in HIV-related risk behaviors is unclear. Few studies have examined changes over time, and the results are inconsistent. For example, in a study of adolescent decision-making regarding contraceptive use, sexually active adolescents visiting health clinics in San Francisco were surveyed in 1984 and 1985 and again in 1985 and 1986.[6] Researchers found no substantial increases in condom use among the participants, and large percentages of respondents continued to report multiple sex partners.

Other studies, however, have found differences over time. To compare risk behavior, one study used data from two matched samples—one from the winter of 1986 and the other from the fall of 1987—of unmarried young adults attending a university in California.[7] Researchers found that students in the later survey not only showed greater worry and concern about all sexually transmitted diseases, including AIDS, but also demonstrated behavioral changes. For example, the proportion of young adults who reported using condoms increased from 46% in 1986 to 62% in 1987.

Respondents surveyed in 1987 also were significantly more likely than those surveyed in 1986 to report delaying sexual involvement, avoiding anal sex, and asking their partners about prior sexual behavior.

Differences also emerged from a Massachusetts study that used data from two independent phone surveys of 16- to 19-year-old adolescents who were selected by statewide random-digit dialing; this study was conducted in August through September of 1986 and 1988.[8] The proportion of teenagers who had a teacher who discussed AIDS in school increased from 52% in 1986 to 82% in 1988. Knowledge about how AIDS is transmitted also was more widespread in 1988. Between 1986 and 1988 some behavioral changes were found as well: reported intravenous drug use declined from 1% to 0.1%; the proportion of sexually active teenagers who reported having made changes in sexual behavior to avoid AIDS increased from 16% to 34%; and the proportion who reported using condoms in response to AIDS increased from 2% to 19%. However, the proportion of teenagers who reported having had sexual intercourse during the year preceding the survey increased from 55% to 61%.

In a retrospective assessment, 39% of inner-city adolescents surveyed in 1988 reported that concern about AIDS led them to make behavior changes during the 6 months preceding the survey.[13] Of those reporting behavior changes, 67% (25% of the total study group) reported that they used condoms, and 16% (6% of the total study group) reported that they were currently abstaining from sexual intercourse.

Although some important behavioral changes were noted in these studies (differences in methodology and samples may partially account for some of the inconsistencies in the results across studies), none represent a very broad cross-section of the adolescent population; therefore, all were unable to assess the impact of the nation's HIV education efforts on behavior. The 1989 Secondary School Student Health Risk Survey and the 1990 national school-based Youth Risk Behavior Survey are the first surveys to measure HIV-related instruction, knowledge, and behaviors among national probability samples of high school students in the United States. Together these surveys provide a unique opportunity to assess possible changes at the national level.

The significantly greater proportion of high school students who in 1990 reported receiving school-based HIV instruction may indicate that more schools are establishing such programs or are expanding current programs to reach students at multiple grade levels. The quality and content of these programs probably vary. Such variation may partly explain our finding that students in the 1990 survey were more knowledgeable about one aspect of HIV transmission but less knowledgeable about another.

Coupled with the increase in school-based HIV instruction, the decreases we found in certain HIV-related behaviors also are encouraging, even though the decreases were not significant for all subgroups of students. These observations are particularly important given the previous trend of increasing sexual activity among selected subgroups of US adolescents over almost two decades.[14,15]

School-based HIV instruction (which appears to be reaching more students over time) and its impact on HIV knowledge may have contributed to some of the positive behavioral changes we observed. Although the survey data do not allow us to make inferences about cause and effect, the similar associations for both surveys suggest that HIV knowledge—which, in turn, was found to be positively associated with HIV education both in the current analysis and in previous analyses[5]—may be one important determinant of HIV-related risk behavior among US high school students. Furthermore, despite the limitations of our knowledge measure in the 1990 survey, the findings from our multivariate analyses held across surveys.

It is important to note that although these findings are suggestive, the relationship between HIV education or knowledge and risk behavior is still unclear. A recent study found an association between AIDS education and (1) increases in the consistency of condom use and (2) decreases in the frequency of intercourse and in the number of sexual partners among teenage men in the United States.[16] AIDS knowledge also was found to be positively correlated with behavior change in an earlier study of inner-city high school students.[13] Other studies of adolescents or young adults, however, have failed to show an association between AIDS-related knowledge and behavioral change or intentions.[8,17-19] Additional research (notably, intervention studies currently under way) may help identify the mechanisms by which education or knowledge can influence behavior in relation to HIV.

It is also important to recognize some limitations of our data. Because behaviors may differ between in-school and out-of-school adolescents, our results can be generalized only to the adolescent population in school. This limitation becomes more pronounced among older youth, because dropout rates increase with student age.

The limitations of self-reported data also apply to our findings; despite extensive measures to ensure student privacy and anonymity, underreporting or overreporting may have occurred. Underreporting by females or overreporting by males may partially explain our finding that males report greater levels of sexual activity than females. Such potential bias, however, is likely to operate in a similar fashion across surveys and therefore is unlikely to influence the changes we observed from 1989 to 1990.

Finally, we must be cautious about interpreting change with only two points in time. Clearly, more data will be needed to identify trends. We can be encouraged, however, by the consistent pattern of decline that emerged for many of these reported risk behaviors—declines that were significant over the 1-year period for some groups of students.

The results of national school-based surveys are useful for assessing our efforts to educate youth about HIV infection. Although limited to two points in time, the behavioral changes we observed between 1989 and 1990 provide information about directions we need to encourage and about those where we need to redouble our efforts. Our findings suggest that school-based HIV instruction may be an important part of the national effort to help us improve the health and well-being of our youth. □

Acknowledgments

Material in this article was presented at the VII International Conference on AIDS, Florence, Italy, July 1991, and the 119th Annual Meeting of the American Public Health Association, Atlanta, Ga, November 1991.

We thank Drs Richard Rothenberg, Luis Escobedo, and Mary Mathis of the Centers for Disease Control and Dr Richard Rubinson of Emory University for their review and thoughtful comments on earlier versions of this manuscript.

References

1. Centers for Disease Control. HIV-related beliefs, knowledge, and behaviors among high school students. *MMWR Morb Mortal Wkly Rep.* 1988;37:717–721.

2. Centers for Disease Control. HIV-related knowledge and behaviors among high school students—selected U.S. sites, 1989. *MMWR Morb Mortal Wkly Rep.* 1990;39: 385–397.
3. Anderson JE, Kann L, Holtzman D, Arday S, Truman B, Kolbe L. HIV/AIDS knowledge and sexual behavior among high school students. *Fam Plann Perspect.* 1990;22:252–255.
4. Kann L, Anderson JE, Holtzman D, Ross J, Truman BI, Collins J, et al. HIV-related knowledge, beliefs, and behaviors among high school students in the United States: results from a national survey. *J Sch Health.* 1991;61:397–401.
5. Holtzman D, Anderson JE, Kann L, Arday SL, Truman BI, Kolbe LJ. HIV instruction, HIV knowledge, and drug injection among high school students in the United States. *Am J Public Health.* 1991;81:1596–1601.
6. Kegeles SM, Adler NE, Irwin CE. Sexually active adolescents and condoms: changes over one year in knowledge, attitudes and use. *Am J Public Health.* 1988;78:460–461.
7. Cochran SD, Keidan J, Kalechstein A. Sexually transmitted diseases and acquired immunodeficiency syndrome (AIDS): changes in risk reduction behaviors among young adults. *Sex Transm Dis.* 1990;17:80–86.
8. Hingson R, Strunin L, Berlin B. Acquired immunodeficiency syndrome transmission: changes in knowledge and behaviors among teenagers, Massachusetts Statewide Surveys, 1986 to 1988. *Pediatrics.* 1990;85:24–29.
9. Kolbe LJ. An epidemiologic surveillance system to monitor the prevalence of youth behaviors that most affect health. *Health Educ.* 1990;21:44–48.
10. Shah BV. *SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data.* Research Triangle Park, NC: Research Triangle Institute; 1987.
11. Centers for Disease Control. Selected behaviors that increase risk for HIV infection among high school students—United States, 1990. *MMWR Morb Mortal Wkly Rep.* 1992;41:231–240.
12. Holtzman D, Greene BZ, Ingraham GC, Daily LA, Demchuk DG, Kolbe LJ. HIV and health education in the United States: a survey of local school district policies and practices. *J Sch Health.* 1992;62:421–427.
13. Goodman E, Cohall AT. Acquired immunodeficiency syndrome and adolescents: knowledge, attitudes, beliefs, and behaviors in a New York City adolescent minority population. *Pediatrics.* 1989;84: 36–42.
14. Centers for Disease Control. Premarital sexual experience among adolescent women—United States, 1970–1988. *MMWR Morb Mortal Wkly Rep.* 1991;39: 929–932.
15. Sonenstein FL, Pleck JH, Ku LC. Sexual activity, condom use and AIDS awareness among adolescent males. *Fam Plann Perspect.* 1989;21:152–158.
16. Ku LC, Sonenstein FL, Pleck JH. The association of AIDS education and sex education with sexual behavior and condom use among teenage men. *Fam Plann Perspect.* 1992;24:100–106.
17. Brown L, Fritz G. Children's knowledge and attitudes about AIDS. *J Am Acad Child Adolesc Psychiatry.* 1988;27:504–508.
18. DiClemente RJ, Forrest KA, Mickler S, et al. College students' knowledge and attitudes about AIDS and changes in HIV-preventive behaviors. *AIDS Educ Prev.* 1990;2:201–212.
19. Keller SE, Bartlett JA, Schleifer SJ, Johnson RL, Pinner E, Delaney B. HIV-relevant sexual behavior among a healthy inner-city heterosexual adolescent population in an endemic area of HIV. *J Adolesc Health.* 1991;12:44–48.

Tobacco Use by Black and White Adolescents: The Validity of Self-Reports

Karl E. Bauman, PhD, and Susan E. Ennett, PhD

ABSTRACT

Objectives. **Previous studies concluded that Black adolescents use tobacco and other drugs less than White adolescents. The Black–White differences typically were attributed to variations in background and life-style. The objective of the research reported in this paper was to determine whether the presumed difference in tobacco use is due to Black–White differences in the validity of self-reports.**

Methods. **We used biochemical measures to compare the validity of self-reports of tobacco use by 1823 Black and White adolescents and to assess the contribution of variation in validity to Black–White differences in reported tobacco use.**

Results. **The sensitivity of Blacks' reports was significantly less than the sensitivity of Whites' reports. The specificity of Whites' reports was significantly less than the specificity of Blacks' reports. Much of the Black–White differences in reports of cigarette smoking and tobacco use were due to Black–White differences in validity.**

Conclusions. **Studies of Black–White differences should adjust for the invalidity of reports or acknowledge that much of the difference may be due to measurement error. (*Am J Public Health.* 1994;84:394–398)**

Introduction

Studies consistently show that Black adolescents are less likely than White adolescents to use drugs.[1-10] This difference typically is attributed to differences in background and life-style. For example, differential vulnerability to models of drug behavior[8] and the substantial impact on youth of Black churches' fundamentalist orientation[9] have been offered as explanations for relatively low drug use by Black adolescents.

All prior studies of Black–White differences in adolescent drug use have relied entirely on self-reports of drug use. Therefore a very different type of explanation for the difference is that Blacks underreport their drug use more than Whites and Whites overreport their drug use more than Blacks. Although this possibility is sometimes acknowledged, it is more often ignored or dismissed as unlikely.[1,9] Only one study has empirically addressed the possibility that Black adolescents might underreport their drug use more than White adolescents. Mensch and Kandel[11] found that among youths who had reported in 1980 that they had used drugs, more Blacks than Whites reported in 1984 that they had never used drugs. The one study of young adults that compared Blacks and Whites concluded that Blacks underreported smoking more than Whites.[12]

In this paper we compare the validity of self-reports of tobacco use by Black and White adolescents and examine the contribution of invalidity to the Black–White difference in self-reported use. We use biochemical indicators as the standards for self-reports, with full recognition that researchers sometimes use self-reports as the standards for biochemical measures and that the biochemical measures we used are not perfect indicators of tobacco use.[13,14]

Methods

The data were gathered for baseline measures to study the influence of mass media campaigns to prevent smoking. Probability samples of households were identified in 10 standard metropolitan statistical areas of the southeastern United States and screened for adolescents aged 12 through 14 years. From April 1, 1985, through October 13, 1985, interviewers attempted to gather data from all eligible adolescents in these households. When more than one adolescent aged 12 through 14 years resided in a household, one was randomly selected to serve as a subject. Of the 2534 adolescent subjects estimated to be eligible for study, 2102 (83%) participated by completing questionnaires in their homes. Subjects averaged 1 hour to complete the self-administered questionnaire and provide biochemical specimens to measure cigarette smoking and tobacco use. Subjects who were not Black or White and subjects who had missing information on any variable were eliminated from these analyses. More detail on the study methodology is available elsewhere.[15]

Karl E. Bauman is with the Department of Health Behavior and Health Education, School of Public Health, University of North Carolina at Chapel Hill. Susan E. Ennett is with the Research Triangle Institute, Research Triangle Park, NC.

Requests for reprints should be sent to Karl E. Bauman, PhD, Department of Health Behavior and Health Education, School of Public Health, Rosenau Hall CB#7400, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599.

This paper was accepted June 28, 1993.

Race was determined by interviewer observation when possible and by interview when necessary. The use of observation as the primary procedure to classify Blacks and Whites was intended to reduce measurement error by minimizing reading and recording errors and by eliminating missing information. In nearly all cases, Whites and Blacks could be readily distinguished by interviewer observation. When the distinction between Blacks and Whites was not obvious the interviewer asked the subject to indicate his or her race, using standard categories. Asking subjects to indicate their race from a range of choices is a widely used research procedure because race determination sometimes is not straightforward. Interviewers received 3 days' training and were closely supervised throughout data collection. We received no reports during our many discussions with interviewers that this procedure produced error.

The percentages of girls among Whites (47.9%) and Blacks (50.6%) were not significantly different ($\chi^2 = 1.1$, $P = .296$, n = 1823). Similarly, the average ages of Whites (mean = 13.1 years) and Blacks (mean = 13.0) were not significantly different ($t = 1.0$, $P = .348$, n = 1823). Parent education was a six-category variable ranging from less than high school graduate to more than 4 years of college education. The education of White and Black parents differed ($\chi^2 = 72.7$, $P = .0001$, n = 1564); White parents were more likely than Black parents to have graduated from high school (81.7% vs 62.8%) and from college (19.0% vs 10.1%).

The primary self-report measure of cigarette smoking was a single questionnaire item that was agreed on for use by investigators funded by the National Cancer Institute[16] and validated by earlier research.[17] The question was "Which statement below BEST describes YOUR current cigarette smoking behavior?" The 13 response categories ranged from "I have never smoked" to "I smoke two packs or more each day." Subjects who said they usually smoked one or more cigarettes per week were considered to be smokers. Supplementary analyses used a question to measure recent smoking ("How long has it been since you last even puffed on a cigarette?"), with eight response categories ranging from "less than 3 hours ago" to "more than 15 hours ago." Smoking within the previous 9 hours was considered recent smoking. The 12 Black subjects and 18 White subjects who reported that they had used marijuana, cigars, or pipes within the previous 9 hours were excluded to preclude their influence on the alveolar carbon monoxide measure described below. Subjects were also asked, "When did you last use chewing tobacco or snuff?" The nine response categories ranged from "never used" to "7 or more days ago." Smokers and any subjects who reported that they had used chewing tobacco or snuff within the previous 3 days were considered to be tobacco users.

Carbon monoxide is a major measurable chemical in tobacco smoke. Each subject provided an alveolar breath sample by taking a deep breath, holding the breath for 20 seconds, blowing half of the breath into the open air, and blowing the remainder of the breath into a bag. The breath was analyzed for level of carbon monoxide, and subjects with levels of 9 ppm or higher were considered to be smokers.[18]

Cotinine is used to indicate nicotine exposure. Each subject deposited 1.5 mL of saliva into a vial. The saliva was frozen on the day of collection, stored until analysis, and analyzed for cotinine by radioimmunoassay. Subjects with cotinine levels of 10 ng/mL or higher were considered to be users of tobacco.[18]

All subjects were told before completing the questionnaires that the biochemical measures would be used to check their reports of cigarette smoking. Earlier studies suggest that such a disclosure enhances the validity of self-reports.[19,20]

Carbon monoxide is used as the standard for self-report of cigarette smoking and cotinine is used as the standard for self-report of tobacco use. Sensitivity is the number of self-reported users per 100 users according to the standard. Specificity is the number of self-reported nonusers per 100 nonusers according to the standard. Underreports and overreports are said to occur when biochemical and self-report measures are incongruent; underreports are self-reports of nonuse when the biochemical measure is positive and overreports are self-reports of use when the biochemical measure is negative. Considerations that must accompany these definitions are addressed in the first part of the discussion section of this paper.

TABLE 1—Self-Reported Cigarette and Tobacco Use by Black and White Adolescents

	% Reporting Use				
	Black (n = 530)	White (n = 1293)	χ^2	*C*	*P*
Cigarettes	.4	4.6	20.8	.11	<.0001
Tobacco	1.7	9.9	36.4	.14	<.0001

Note. *C* = contingency coefficient.

Results

As other studies have shown, Whites are more likely than Blacks to report that they smoke cigarettes and use tobacco (Table 1).

Table 2 shows the sensitivities and specificities for self-reports of cigarette smoking and tobacco use. Blacks were more likely than Whites to underreport use of these substances. Whites were more likely than Blacks to overreport use. The Black–White difference in sensitivity is much larger than the Black–White difference in specificity. Because the vast majority of subjects are not users, however, the small Black–White difference in specificity can contribute to the Black–White difference in self-reported use. The greater use reported by Whites than by Blacks appears to be influenced both by Black underreports and by White overreports.

Do Black–White differences in validity of reports contribute much to the Black–White differences in smoking and tobacco use? To address this question directly, we compared the ratios of White-to-Black use from self-report and biochemical measures. For this comparison we assumed that most if not all of the difference in ratios is attributable to the invalidity of self-reports. Table 3 shows the percentages of cigarette and tobacco users according to self-report and biochemical measures for Blacks and

TABLE 2—Sensitivity and Specificity of Black and White Adolescents' Self-Reports of Cigarette and Tobacco Use

	Black	White	C	P[a]
Cigarettes				
Sensitivity, %	0	62.5	.37	.002
Reported use, no.	0	35		
CO ≥ 9 ppm, no.	7	56		
Specificity, %	99.6	98.0	.06	.009
Reported nonuse, no.	521	1212		
CO < 9 ppm, no.	523	1237		
Tobacco				
Sensitivity, %	12.5	55.8	.25	.001
Reported use, no.	2	82		
Cotinine ≥ 10 ng/ml, no.	16	147		
Specificity, %	98.6	96.0	.07	.004
Reported nonuse, no.	507	1100		
Cotinine < 10 ng/ml, no.	514	1146		

Note. C = contingency coefficient; CO = carbon monoxide.
[a]From Fisher's Exact Test.

TABLE 3—Unadjusted Rates of Tobacco Use by Black and White Adolescents According to Self-Report and Biochemical Measures

	Use, %		Ratio of White to Black Use
	White	Black	
Cigarettes			
Self-report	4.6	.4	11.5
Carbon monoxide	4.3	1.3	3.3
Tobacco			
Self-report	9.9	1.7	5.8
Continine	11.4	3.0	3.8

Whites. The ratios of the White to Black percentages are shown in the right column. Of particular interest is that the ratios of White to Black use are substantially smaller for biochemical measures than for self-reports, suggesting that Black–White differences in cigarette and tobacco use are substantially increased by Black–White differences in invalid self-reports.

Another way to describe the contribution of invalidity to the difference in Black and White-self-reports is to first adjust the self-reports of Blacks for the sensitivity and specificity of Whites and then compare the White-to-Black ratio of unadjusted rates to the ratio of White unadjusted to Black adjusted rates. The extent to which the first ratio is larger than the second reflects the contribution of invalidity to Black–White differences in self-reports. Parallel comparisons can be made with White self-reports adjusted for Black sensitivity and specificity.

The procedures for calculating the adjusted rates are as follows. Sensitivity and specificity, expressed as percentages in Table 2, are converted to proportions by dividing by 100. To derive the number of Blacks who smoke under the conditions of White sensitivity and specificity, we summed two products: (1) White sensitivity times the number of Blacks with carbon monoxide levels of 9 ppm or higher ($.625 \times 7 = 4.375$) and (2) 1.00 minus white specificity ($1.00 - 0.98 = 0.02$) times the number of Blacks with carbon monoxide levels of less than 9 ppm ($0.02 \times 523 = 10.460$). The adjusted percentage of self-reported Black smokers is the above sum ($4.375 + 10.460 = 14.835$) per 100 Black subjects ($14.835/530 \times 100 = 2.8$). The same procedure was used to derive a rate of 5.6% of Black self-reported tobacco users adjusted for the sensitivity and specificity of Whites. We used the same procedure to derive rates for Whites that were adjusted for Black sensitivity and specificity. The percentages of White self-reported cigarette and tobacco users, adjusted for Black sensitivity and specificity, are 0.4 and 2.7, respectively.

The unadjusted White-to-Black ratio of 11.5 for cigarette use (Table 3) is much larger than the 1.6 ratio of White unadjusted to Black adjusted rates (4.6:2.8) and the 1.0 ratio of White adjusted to Black unadjusted rates (0.4:0.4). For tobacco use, the unadjusted White-to-Black ratio of 5.8 is substantially larger than the 1.8 ratio of White unadjusted to Black adjusted rates (9.9:5.6) and the 1.6 ratio of White adjusted to Black unadjusted rates (2.7:1.7). These comparisons, like those presented earlier that compared the ratios of White to Black use from self-report and biochemical measures (Table 3), suggest that a substantial portion of the Black–White difference in self-reported use is due to Black–White differences in the validity of self-reports.

The findings presented above use smoking within a week to identify self-reported smokers. Because carbon monoxide can be detected for about 9 hours, reports of smoking within the previous 9 hours might be more meaningful for determining sensitivity and specificity. However, sensitivity and specificity are changed only slightly when recency is substituted for smoking. Specifically, for Blacks, sensitivity for smoking and tobacco use were identical when the different self-reports of cigarette smoking were used, and specificity was 2.5% lower with the recency measure. When recency was substituted for smoking within a week for Whites, sensitivity of smoking decreased from 62.5% to 53.6% and sensitivity of tobacco use decreased from 55.8% to 50.3%; specificity decreased by 0.2% and 1.5% for smoking and tobacco use, respectively.

Discussion

In this paper we treat biochemical indicators as standards for self-reports of smoking and tobacco use. There are, however, no gold standards for these behaviors. Recent comprehensive reviews of the power, limitations, and range of interpretation of biochemical measures of smoking are available elsewhere.[13,14] Neither of the biochemical indicators has a point at which self-reported behaviors can be determined precisely. The amount of time carbon monoxide and cotinine remain detectable in body fluids varies somewhat by various characteristics,[18] and carbon monoxide level can be influenced by factors such as ambient tobacco smoke and automobile exhaust. We set our

cut-point for carbon monoxide where contribution from sources other than active smoking is unlikely, and the cut-points for both biochemical measures are those commonly used by other researchers. Different self-report measures of cigarette smoking produced sensitivity virtually identical to the values in our tables. Extensive research has failed to find substances other than tobacco that produce cotinine levels as high as those we used to indicate tobacco use, but such substances may be discovered in the future.

It is possible that Blacks and Whites have such large differences in carbon monoxide and cotinine levels for reasons other than smoking and tobacco use that those reasons, rather than the invalidity of self-reports, explain the Black–White difference in biochemical and self-report congruency. If this is the case, then what we consider to be Black–White differences in the validity of self-reports in our data are not differences in invalidity at all. Metabolism of nicotine or excretion of cotinine may vary by race in young adults and therefore in young adolescents, and may thereby influence estimates of sensitivity and specificity.[21] We assume that such factors are too weak to entirely explain the large Black–White differences in sensitivity in our data and the stability of estimates across self-report and biochemical measures. Until future research clarifies this point, however, it must be recognized that what appear to be large Black–White differences in the validity of self-reports may actually reflect something else.

We would have preferred to have more Black smokers and tobacco users available for estimating the sensitivity of self-reports. Indeed, of the 530 Black subjects, only 2 reported smoking and 7 reported tobacco use. Even with the small numerators, however, the differences between Blacks and Whites were statistically significant.

Invalidity may contribute even more to the difference between Black and White self-reports in studies that do not inform subjects that their self-reports will be compared with biochemical indicators. It is likely that in our study validity was increased, and perhaps the difference in invalidity between Blacks and Whites was reduced, because our subjects were told that their answers would be checked by the biochemical indicators.[19,20]

An advantage of the data used in this study is that they did not depend on school enrollment. Most studies of adolescent drug use are based on data collected in schools. The higher school dropout rate of Blacks and the high use of drugs reported by dropouts[22] and students with high absenteeism[23] complicate Black–White comparisons when school-based data are used.

Others have argued that self-reports obtained from adolescents in their homes are less valid than those gathered in schools.[23] The extent to which this is true could limit the generalizability of our findings because our data were collected in homes and many studies of adolescents use data collected in schools. However, the only examination of this possibility that used a true experimental design and thereby avoided many of the potential confounds, such as markedly different sample frames and instrumentation, found no difference in drug use between data gathered in homes and in schools.[24]

Our findings may not be generalizable to drugs other than tobacco. Drug behaviors are significantly correlated, however, and there are compelling reasons for subjects to underreport their use of other drugs in addition to tobacco. Our findings indicate that Black–White differences in invalidity should not be readily dismissed when Black–White comparisons are made.

Others have identified variables that need to be taken into account when attempting to explain Black–White differences in adolescent tobacco use. They include religion, vulnerability, parents' and friends' use, availability, and risk taking.[5,8,9] Our findings suggest that differential invalidity is one of the more important explanatory variables.

Conclusions

Future studies of Black–White differences in drug use that rely on self-reports should account for invalid measurement before proceeding to examine variables of more theoretical and practical interest, or they should give more credence to the possibility that the differences they attempt to explain may be due in large part to differential validity. □

Acknowledgments

This research was funded by grant CA38392 from the National Cancer Institute of the National Institutes of Health.

Thanks to Dr Nancy J. Haley for laboratory analyses of cotinine and related advice and to Dr Mario Orlandi for his contribution to the availability of the cotinine measures.

References

1. Bachman JG, Wallace JMJ, O'Malley PM, Johnston LD, Kurth CL, Neighbors HW. Racial/ethnic differences in smoking, drinking, and illicit drug use among American high school seniors, 1976–89. *Am J Public Health.* 1991;81:372–377.
2. Bauman KE, Koch GG, Fisher LA, Bryan ES. Use of smokeless tobacco by age, race, and gender in ten standard metropolitan statistical areas of the Southeast United States. *Natl Cancer Inst Monogr.* 1989;8:35–37.
3. Centers for Disease Control. Cigarette smoking among youth—United States, 1989. *MMWR.* 1991;40:712–715.
4. Harford TC. Drinking patterns among black and nonblack adolescents: results of a national survey. *Ann N Y Acad Sci.* 1986;472:130–141.
5. Headen SW, Bauman KE, Deane GD, Koch GG. Are the correlates of cigarette smoking initiation different for black and white adolescents? *Am J Public Health.* 1991;81:854–858.
6. Kandel D, Single E, Kessler RC. The epidemiology of drug use among New York state high school students: distribution, trends, and change in rates of use. *Am J Public Health.* 1976;66:43–53.
7. Moss AJ, Allen KF, Giovino GA, Mills SL. *Recent Trends in Adolescent Smoking, Smoking-Uptake Correlates, and Expectations About the Future.* Hyattsville, Md: National Center for Health Statistics; 1992. DHHS publication 93-1250.
8. Newcomb MD, Bentler PM. Substance use and ethnicity: differential impact of peer and adult models. *J Psychol.* 1986;120:83–95.
9. Wallace JMJ, Bachman JG. Explaining racial/ethnic differences in adolescent drug use: the impact of background and lifestyle. *Soc Probl.* 1991;38:333–357.
10. Welte JW, Barnes GM. Alcohol use among adolescent minority groups. *J Stud Alcohol.* 1987;48:329–336.
11. Mensch BS, Kandel DB. Underreporting of substance use in a national longitudinal youth cohort: individual and interviewer effects. *Public Opin Q.* 1988;52:100–124.
12. Wagenknecht LE, Burke GL, Perkins LL, Haley NJ, Friedman GD. Misclassification of smoking status in the CARDIA study: a comparison of self-report with serum cotinine levels. *Am J Public Health.* 1992;82:33–36.
13. Dolcini MM, Adler NE, Ginsberg D. When can you trust what they say? A review of the validity of self-reported smoking among adolescents and pregnant women. San Francisco, Calif: Division of General Internal Medicine and Center for AIDS Prevention Studies, University of California, San Francisco; 1993. Unpublished manuscript.
14. Velicer W, Prochaska J, Rossi J, Snow M. Assessing outcome in smoking cessation studies. *Psych Bull.* 1992;111:23–42.
15. Bauman KE, LaPrelle J, Brown JD, Koch

GG, Padgett CA. The influence of three mass media campaigns on variables related to adolescent cigarette smoking: results of a field experiment. *Am J Public Health.* 1991;81:597–604.

16. *Smoking, Tobacco, and Cancer Program: 1985 Report.* Bethesda, Md: National Cancer Institute; 1986. DHHS publication 86-2687.
17. Pechacek TF, Murray DM, Luepker RV, Mittelmark MB, Johnson CA, Shutz JM. Measurement of adolescent smoking behavior: rationale and methods. *J Behav Med.* 1984;7:123–140.
18. Bauman KE, Koch GG, Bryan ES, Haley NJ, Downton MI, Orlandi MA. On the measurement of tobacco use by adolescents: validity of self-reports of smokeless tobacco use and validity of cotinine as an indicator of cigarette smoking. *Am J Epidemiol.* 1989;130:327–337.
19. Bauman KE, Dent CW. Influence of an objective measure on self-reports of behavior. *J Appl Psychol.* 1982;67:623–628.
20. Murray DM, O'Connell CM, Schmid LA, Perry CL. The validity of smoking self-reports by adolescents: a reexamination of the bogus pipeline procedure. *Addict Behav.* 1987;12:7–15.
21. Wagenknecht LE, Cutter GR, Haley NJ, et al. Racial differences in serum cotinine levels among smokers in the coronary artery risk development in (young) adults study. *Am J Public Health.* 1990;80:1053–1056.
22. Pirie PL, Murray DM, Luepker RV. Smoking prevalence in a cohort of adolescents, including absentees, dropouts, and transfers. *Am J Public Health.* 1988;78:176–178.
23. Kandel DB. *Longitudinal Research on Drug Use: Empirical Findings and Methodological Issues.* Washington, DC: Hemisphere-Wiley; 1978.
24. Zanes A, Matsoukas E. Different settings, different results? A comparison of school and home responses. *Public Opin Q.* 1979;43:550–557.

Smokeless Tobacco Use and Increased Cardiovascular Mortality among Swedish Construction Workers

Gunilla Bolinder, MD, Lars Alfredsson, PhD, Anders Englund, MD, and Ulf de Faire, MD, PhD

ABSTRACT

Objectives. Little is known about the risks of cardiovascular disease associated with the use of smokeless tobacco, which produces blood nicotine levels similar to those caused by cigarette smoking.

Methods. Male Swedish construction industry employees (n = 135 036) who attended a health examination were followed by studying cause-specific mortality during a 12-year period. The study population comprised 6297 smokeless tobacco users, 14 983 smokers of fewer than 15 cigarettes per day, 13 518 smokers of 15 or more cigarettes per day, 17 437 ex-smokers, 50 255 "other" tobacco users, and 32 546 nonusers.

Results. The age-adjusted relative risk of dying from cardiovascular disease was 1.4 for smokeless tobacco users and 1.9 for smokers of 15 or more cigarettes per day, compared with nonusers. Among men aged 35 through 54 years at the start of follow-up, the relative risk was 2.1 for smokeless tobacco users and 3.2 for smokers. When data were adjusted for body mass index, blood pressure, and history of heart symptoms, the results were essentially unchanged. Cancer mortality was not raised in smokeless tobacco users.

Conclusions. Both smokeless tobacco users and smokers face a higher risk of dying from cardiovascular disease than nonusers. Although the risk is lower for smokeless tobacco users than for smokers, the excess risk gives cause for preventive actions. (*Am J Public Health.* 1994;84:399–404)

Introduction

Although there is well-established evidence for a causal link between cigarette smoking and cardiovascular damage and disease,[1] the reasons behind the development of atherosclerosis in smokers are not fully understood. The nicotine content of cigarette smoke is a suspected contributor to vascular damage, as nicotine has a variety of potentially relevant cardiovascular effects. Nicotine most likely promotes atherosclerotic disease by its actions on lipid metabolism and coagulation,[2–5] by hemodynamic effects,[1,4] and/or by causing endothelial injury.[6]

However, there have been no studies on long-term exposure to pure nicotine in humans, and only limited experience from animal studies.[7] Therefore, users of smokeless tobacco (snuff) present an excellent opportunity for studying the effects of exposure to nicotine without simultaneous exposure to carbon monoxide and other combustion products of tobacco smoke. It has been shown that the use of smokeless tobacco, during which nicotine is absorbed through the buccal mucosa, produces maximum blood levels of nicotine similar to those produced by cigarette smoking and results in a larger overall exposure to nicotine owing to prolonged absorption.[8,9]

In a recent cross-sectional study, it was observed that middle-aged and older smokeless tobacco users had a significantly higher prevalence of hypertension than either nonusers or smokers.[10] The smokeless tobacco users were also affected more than nonusers by cardiovascular symptoms. During a 4-year follow-up period, smokeless tobacco users had a higher risk than nonusers of disability retirement due to cardiovascular diseases. The excess risk was quite similar to the excess risk observed in smokers.

The aim of the present study was to investigate whether long-term exposure to smokeless tobacco is associated with an excess risk of dying from cardiovascular disease in users compared with nonusers and to compare this potential excess risk among smokeless tobacco users with the corresponding excess risk among cigarette smokers.

Materials and Methods

Subjects

The study population comprised 135 036 men who had received medical checkups under the auspices of the Swedish Construction Industry's Organization for Working Environment Safety and Health during the years 1971 through 1974. The cohort included construction workers, electricians, painters, sheet metal workers, and other construction industry employees in the whole of Sweden. The invitation to the voluntary medical examination was sent out by the construction site staff and about 75%

Gunilla Bolinder and Ulf de Faire are with the Department of Medicine, Karolinska Hospital, Stockholm, Sweden. Lars Alfredsson is with the Division of Epidemiology, Institute of Environmental Medicine, Karolinska Institute, Stockholm. Anders Englund is with the Swedish Construction Industry's Organization for Working Environment Safety and Health, Danderyd, Sweden.

Requests for reprints should be sent to Gunilla Bolinder, MD, Department of Medicine, Karolinska Hospital, S-17176 Stockholm, Sweden.

This paper was accepted June 28, 1993.

TABLE 1—Distribution of the Study Population, by Age and Tobacco Habit

Age	Nonusers	Smokeless Tobacco Users	Smokers, <15 cig/d	Smokers, ≥15 cig/d	Ex-Smokers	Total
<35 y	13 120	2891	7 626	6 350	4 369	34 356
35–54 y	13 784	1672	5 225	5 785	8 222	34 688
55–65 y	5 642	1734	2 132	1 383	4 846	15 737
All	32 546	6297	14 983	13 518	17 437	84 781

Note. Those who had mixed tobacco use or who smoked pipes or cigars (50 255 subjects, or 37.2% of the total sample of 135 036) are excluded from table.

responded to it. In 1974 there were 183 865 registered workers in the Swedish construction industry; it is not known to what extent the nonattending workers (≈25%) were not reached by the invitation or were not willing to come. Women (< 0.5% of workers in the construction industry) were excluded from the study.

Procedures

A standardized program was followed for the medical examination. Body weight and height were determined. Blood pressure was measured after 5 minutes of rest in a supine position. A standard mercury manometer on the upper arm was used to record the systolic level when phase 1 (Korotkoff) sounds were heard and the diastolic level at phase 5 (disappearance of sounds). Blood pressure was recorded to the nearest 2 mm Hg. Heart rate was recorded as beats per minute, simultaneously with the blood pressure measurement.

Tobacco habits by kind of tobacco use (cigarettes, smokeless tobacco, pipe, cigar, or mixed use), amount of smoked tobacco, and duration of the tobacco consumption habit were recorded with the aid of a questionnaire that the subject filled out together with a nurse. The questionnaire also requested information about present or past history of symptoms and diseases as well as current medication.

The members of the study population identified between 1971 and 1974 and who were alive on January 1, 1974, were followed regarding cause-specific mortality during the period 1974 through 1985 with the aid of the National Cause of Death Register,[11] with certification of the underlying cause of death according to the *International Classification of Diseases*, 8th revision (ICD-8). The underlying cause of death is defined as the illness that started the chain of events that directly led to death. The diagnoses specifically studied were ischemic heart disease (ICD 410–414), cerebrovascular disorders (ICD 430–438), all cardiovascular diagnoses (ICD 390–458), and all malignant neoplasms (ICD 140–209).

Analyses

The classification of tobacco habits was aimed at isolating subjects in groups with a single type of exposure to tobacco. Tobacco users were divided into smokeless tobacco users, smokers, ex-smokers, and others. Nonusers were subjects who reported that they had never used tobacco. Smokers were divided into cigarette smokers who smoked fewer than 15 cigarettes per day and those who smoked 15 cigarettes per day or more, both groups being without any other kind of former or present tobacco use. Smokeless tobacco users were subjects who reported only present smokeless tobacco use and no former or present smoking. Ex-smokers were cigarette smokers who had never smoked a pipe, smoked cigars, or used smokeless tobacco and who had quit smoking; they were further divided into those who had not smoked for less than 5 years and those who had not smoked for 5 or more years. "Others" were all subjects with mixed tobacco use or subjects who smoked a pipe or cigars. The duration of the tobacco habit was divided into less than 15 years and 15 or more years. Because smokers of more than 15 cigarettes per day consume about the same amount of nicotine as daily smokeless tobacco users,[9,12] special emphasis should be placed on this group when interpreting the analysis.

In the analysis, blood pressure was divided into four categories: systolic blood pressure ≤ 140, 141–159, 160–179, and ≥180 mm Hg and diastolic blood pressure <85, 85–94, 95–104, and ≥105 mm Hg. Blood pressure and heart rate data were missing for 0.2% of the subjects. The body mass index was calculated as body weight in kilograms divided by height in meters squared. Body mass indexes were divided into four categories: ≤20.0, 20.1–25.0, 25.1–30.0, and >30.0.

Relative risks of death due to specific causes, together with 95% confidence intervals, were estimated for subjects with different tobacco habits in comparison with the nonusers. To adjust for potential confounding factors, the Mantel-Haenszel procedure was used[13] and 95% confidence intervals were estimated according to Greenland et al.[14] Potential confounding factors were considered to be age (5-year brackets), area of domicile, blood pressure, blood pressure medication, previous cardiac symptoms, diabetes, and body mass index.

Analyses of the relationship between different tobacco habits and mortality were performed for the entire cohort, for workers entering the study at ages 35 through 54 years, and for those entering the study at ages 55 through 65 years.

Results

The distribution of the study population by tobacco habit is shown in Table 1; 24.1% of the subjects were nonusers, 4.7% were smokeless tobacco users, 11.1% were smokers of fewer than 15 cigarettes per day, 10.0% were smokers of more than 15 cigarettes per day, and 12.9% were ex-smokers. The 37.2% of the population who smoked pipes or cigars or who had mixed tobacco use (n = 50 255) were excluded from all analyses.

During the follow-up period (1974 through 1985), there were a total of 8293 deaths in the whole study cohort. Fifty-seven percent of the deaths occurred in the analyzed tobacco-use groups: nonusers, 1322; smokeless tobacco users, 440; smokers of fewer than 15 cigarettes per day, 900; smokers of 15 or more cigarettes per day, 923; and ex-smokers, 1126. Cardiovascular disease was the most common cause of death among the construction workers. Ischemic heart disease caused 38% of the deaths in the whole study population, implying a crude cumulative mortality of 2.3% during the 12-year follow-up period.

TABLE 2—Observed Numbers of Deaths and Relative Risks (RRs)[a] for Causes of Death for Different Tobacco Habit Groups Compared with Nonusers

Cause of Death	Nonusers[b] (At-Risk n = 32 546)	Smokeless Tobacco Users (At-Risk n = 6297)		Smokers, <15 cig/d (At-Risk n = 14 983)		Smokers ≥15 cig/d (At-Risk n = 13 518)		Ex-Smokers, 1–5 y (At-Risk n = 6761)		Ex-Smokers, >5 y (At-Risk n = 9800)	
	No.	No.	RR (95% CI)	No.	RR (95% CI)	No.	RR (95% CI)	No.	RR (95% CI)	No.	RR (95% CI)
All cardiovascular disease	641	220	1.4 (1.2, 1.6)	450	1.8 (1.6, 2.0)	381	1.9 (1.7, 2.2)	169	1.4 (1.1, 1.6)	402	1.1 (0.9, 1.2)
All cancer	372	96	1.1 (0.9, 1.4)	216	1.5 (1.3, 1.8)	276	2.5 (2.2, 3.0)	119	1.6 (1.3, 2.0)	249	1.3 (1.1, 1.6)
All causes	1322	440	1.4 (1.3, 1.8)	900	1.7 (1.6, 1.9)	923	2.2 (2.0, 2.4)	350	1.3 (1.2, 1.5)	776	1.1 (1.0, 1.2)

Note. The deaths of 3582 men who were pipe or cigar smokers or had a mixed tobacco use habit are not presented in the table. For 876 ex-smokers (5%), length of time since quitting smoking was unknown; these ex-smokers were not included in this analysis. CI = confidence interval.
[a]Adjusted for age in 5-year intervals and for region of origin.
[b]Nonusers are the referent category (RR = 1.0).

Table 2 shows the age-adjusted relative risks of dying of cardiovascular disease, malignant tumors, and any cause for different tobacco-use groups compared with the nonusers. The excess risk of dying of cardiovascular disease was most pronounced for smokers, with a dose–response relation. In ex-smokers, the excess risk diminished with the time since smoking was stopped. Smokeless tobacco use was also found to be associated with an excess risk of dying of cardiovascular disease, whereas regarding death from cancer, no excess risk was observed. Although the present study did not analyze different cancer diagnoses in detail, it was obvious that smoking had a dose–response relationship to overall cancer risk and that the excess risk diminished gradually with the time since smoking was stopped.

Table 3 shows the results of the analysis of more specific causes of death in workers in two age groups. Higher relative risks were observed in the younger than in the older group for both smokeless tobacco users and smokers, compared with the nonusers. For smokeless tobacco users compared with nonusers, the relative risk of dying of cardiovascular disease was 2.1 (95% confidence interval [CI] = 1.5, 2.9) for those in the younger group and only 1.1 (95% CI = 1.0, 1.4) for those in the older group.

Death from stroke was less common than death from ischemic heart disease, but with regard to tobacco habits, the risk patterns for ischemic heart disease and stroke were the same. The relative risk of stroke among the younger smokeless tobacco users was close to 2, compared with the nonusers, but the number of cases was small and therefore the confidence interval was wide.

Lung cancer was studied specifically to evaluate whether hidden smokers could be found among the declared smokeless tobacco users. Three deaths from lung cancer were found in this group, whose relative risk compared with the nonusers was 0.9 (95% CI = 0.2, 3.0).

When potential confounding due to age, area of domicile, body mass index, blood pressure, diabetes, and history of heart symptoms or blood pressure medication at the time of entering the study was analyzed according to the Mantel-Haenszel procedure, the relative risks of death from cardiovascular diseases remained essentially unchanged. For cancer and overall mortality, no changes in the relative risk estimates were found when confounding factors were considered.

It was not considered meaningful to evaluate the duration of the tobacco habit in relation to cardiovascular mortality because very few subjects in age groups prone to various cardiovascular manifestations (i.e., those older than 45 years) exhibited a duration of tobacco use of less than 15 years, as most tobacco users start the habit at a young age. Eighty-seven percent of the cardiovascular deaths among smokeless tobacco users were associated with a tobacco habit duration of more than 15 years at the time the subjects entered the study. The corresponding figure for smokers of fewer than 15 cigarettes per day was 89%; for smokers of 15 or more cigarettes per day, the figure was 97%.

Figure 1 shows, for subjects who entered the study at ages 35 through 54 years, the relative risk of dying of ischemic heart disease for subjects in different tobacco-use categories compared with the nonusers. For smokeless tobacco users, the risk of death from ischemic heart disease was higher than that of nonusers and ex-smokers but lower than that of cigarette smokers.

Discussion

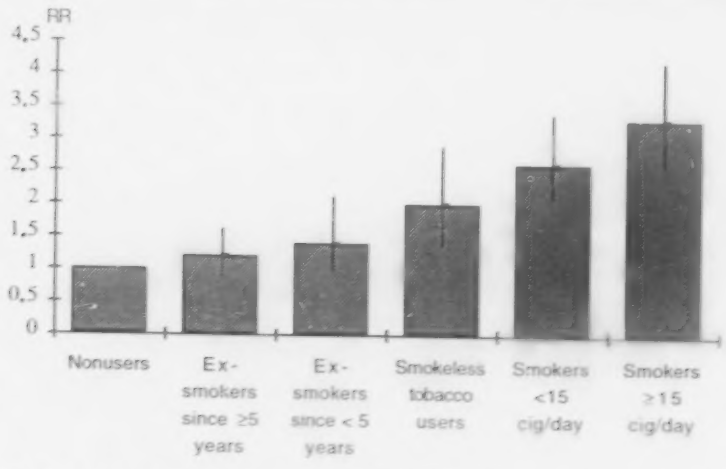
The present study of cardiovascular and cerebrovascular mortality in a cohort comprising more than 6000 smokeless tobacco users is the first observational epidemiological cohort study with the possibility of comparing a large group of smokeless tobacco users with both smokers and subjects not using any kind of tobacco. The results indicate an apparent excess risk of death from cardiovascular and cerebrovascular diseases of about 40% to 100% among smokeless tobacco users, compared with nonusers, when possible confounding factors are taken into account. Smokers face even higher risks of both cardiovascular and cerebrovascular causes of death.

In a recent case–control study from northern Sweden encompassing 35- through 64-year-old men with a first myocardial infarction and population-derived controls, smoking but not the use of smokeless tobacco was associated with an increased risk of myocardial infarction.[15] These results, which are contradictory to ours, could be due to differences in study design and composition of reference groups. Reasonably, the power to detect potential hazard effects from smokeless tobacco use is greater in an observational cohort study

TABLE 3—Cause-Specific Mortality among Swedish Construction Workers during 12-Year Follow-Up after 1971–1974 Health Examination

Cause of Death	Nonusers[a]	Smokeless Tobacco Users		Smokers, <15 cig/d		Smokers ≥15 cig/d		Ex-Smokers, 1–5 y		Ex-Smokers, >5 y	
	No.	No.	RR (95% CI)	No.	RR (95% CI)	No.	RR (95% CI)	No.	RR (95% CI)	No.	RR (95% CI)
Age 35–54 years at entry into study											
	(At-Risk n = 13 784)	(At-Risk n = 1672)		(At-Risk n = 5225)		(At-Risk n = 5785)		(At-Risk n = 2882)		(At-Risk n = 5005)	
Ischemic heart disease	123	35	2.0 (1.4, 2.9)	128	2.6 (2.1, 3.4)	162	3.3 (2.6, 4.2)	37	1.4 (1.0, 2.1)	67	1.2 (0.9, 1.6)
Stroke	16	4	1.9 (0.6, 5.7)	17	2.7 (1.4, 5.4)	19	3.0 (1.5, 5.7)	4	1.2 (0.4, 3.7)	5	0.7 (0.2, 1.9)
All cardiovascular disease	154	44	2.1 (1.5, 2.9)	164	2.7 (2.2, 3.4)	199	3.2 (2.6, 3.9)	46	1.4 (1.0, 2.0)	83	1.1 (0.9, 1.5)
All cancer	128	22	1.2 (0.8, 1.9)	62	1.2 (0.9, 1.7)	116	2.2 (1.8, 2.9)	42	1.6 (1.1, 2.2)	62	1.1 (0.8, 1.5)
Lung cancer	5	1	1.2 (0.2, 9.1)	16	8.1 (3.2, 20.4)	43	21.4 (8.5, 54.1)	7	6.7 (2.3, 19.7)	3	1.2 (0.3, 4.5)
All causes	410	105	1.9 (1.6, 2.4)	317	2.0 (1.7, 2.3)	437	2.6 (2.3, 3.0)	114	1.3 (1.1, 1.6)	189	1.0 (0.9, 1.2)
Age 55–65 years at entry into study											
	(At-Risk n = 5642)	(At-Risk n = 1734)		(At-Risk n = 2132)		(At-Risk n = 1383)		(At-Risk n = 1076)		(At-Risk n = 3660)	
Ischemic heart disease	359	137	1.2 (1.0, 1.5)	225	1.7 (1.4, 1.9)	122	1.4 (1.2, 1.8)	89	1.3 (1.1, 1.6)	248	1.1 (0.9, 1.2)
Stroke	70	26	1.2 (0.7, 1.8)	19	0.7 (0.4, 1.2)	25	1.6 (1.0, 2.5)	20	1.5 (0.9, 2.5)	35	0.8 (0.5, 1.2)
All cardiovascular disease	480	174	1.1 (1.0, 1.4)	272	1.5 (1.3, 1.7)	167	1.5 (1.3, 1.7)	120	1.3 (1.1, 1.6)	317	1.0 (0.9, 1.2)
All cancer	223	69	1.0 (0.8, 1.3)	145	1.7 (1.4, 2.1)	148	2.9 (2.3, 3.5)	69	1.6 (1.3, 2.1)	183	1.3 (1.1, 1.5)
Lung cancer	8	2	0.8 (0.1, 3.9)	36	11.9 (5.5, 25.6)	57	30.6 (14.6, 64.1)	14	9.4 (3.9, 22.3)	12	2.3 (1.0, 5.7)
All deaths	820	301	1.2 (1.0, 1.3)	496	1.6 (1.5, 1.8)	377	2.0 (1.8, 2.2)	212	1.4 (1.2, 1.6)	576	1.1 (1.0, 1.2)

Note. For some ex-smokers (335 [4.1%] of those aged 35–54 years at entry and 110 [2.3%] of those aged 55–65 years at entry), length of time since quitting smoking was unknown; these ex-smokers were not included in this analysis. No. = number of deaths in the group; RR = relative risk, adjusted for age (5-year intervals) and for region of origin, for tobacco users compared with nonusers; CI = confidence interval.
[a]Nonusers are the referent category (RR = 1.0).



Note. Only subjects who were 35–54 years of age at entry into the study are included.

FIGURE 1—Adjusted relative risk (RR), together with 95% confidence interval, of dying due to ischemic heart disease among subjects in different tobacco habit groups as compared with nonusers.

with 3159 deaths due to ischemic heart disease (of which 172 were smokeless tobacco users) than in a smaller case-control study with 585 incident cases of myocardial infarction (of which 59 were smokeless tobacco users). In addition, our nonusers had never used any tobacco, whereas the nonusers in the case-control study included both former smokers and former smokeless tobacco users. The use of such "nonusers" might have contributed to a dilution of a possible excess risk for myocardial infarction in smokeless tobacco users.

Possible bias due to misclassification of both exposure and disease should be considered in our study. Assuming that smokers have a higher risk of dying from cardiovascular disease than smokeless tobacco users, a misclassification of smokers in the smokeless tobacco users group would lead to an overestimation of the relative risk associated with smokeless tobacco (i.e., the observed relative risk would be too high). How-

ever, if the proportion of misclassified smokers in the group classified as using smokeless tobacco was considerable, an elevated risk of dying from lung cancer would have been expected. Only three deaths from lung cancer were found among those classified as smokeless tobacco users. The corresponding relative risk of dying from lung cancer, compared with nonusers, was 0.9 (95% CI = 0.2, 3.0). In addition, different kinds of smokers or smokeless tobacco users wrongly classified as nonusers would result in an underestimation of the correlation between smokeless tobacco use and mortality due to cardiovascular disease. On combining different sources of bias due to misclassification of exposure, one finds that an underestimation of the relative risk of death from cardiovascular disease associated with smokeless tobacco use is more likely than an overestimation.

No follow-up recording of subjects' tobacco habits after entry into the study has been made. The most common change in tobacco habits in adults is to quit using tobacco. If our subjects had quit using tobacco, the relative risk estimates might have been diluted compared with what would be expected if no change took place during the follow-up period. If smokeless tobacco users had a tendency to start smoking, the estimated relative risks for smokeless tobacco users would be exaggerated. Nearly all smokeless tobacco users had used smokeless tobacco for more than 15 years without smoking, so it is reasonable to assume that the possible influence of such changes has been small.

The completeness of the Swedish National Cause of Death Register is almost 100%, as all deaths are required to be registered.[16] The validity of the underlying causes of death stated in death certificates has been evaluated and found to be high regarding cardiovascular diseases, stroke, and cancer.[17] Possible misclassifications are not likely to be dependent on the type of tobacco habit and therefore should not affect the relative risk estimates.[18]

The relationship between smokeless tobacco use and mortality due to specific causes was scrutinized regarding potential confounding from age, area of domicile, body mass index, blood pressure, diabetes, and history of heart symptoms or blood pressure medication at the time of entering the study. Cardiovascular risk markers such as total cholesterol were not routinely measured at the time of the medical examinations of this cohort, and therefore it was not possible to adjust for hypercholesterolemia as a confounding factor in the analysis. Cholesterol and high-density lipoprotein cholesterol did not differ between smokeless tobacco users and nonusers in two studies of relatively young users of smokeless tobacco,[19,20] but significantly higher levels of plasma cholesterol were found in smokeless tobacco users in one study in which older users were also examined.[21] The body mass index of smokeless tobacco users is generally higher than that of nonusers and considerably higher than that of smokers,[10] possibly implying differences in lipoprotein levels. However, the body mass index regarded as a confounder in the analysis did not change the relative risk estimates for the association between smokeless tobacco use and mortality due to cardiovascular disease.

It was not possible to consider potential confounding due to alcohol use, as no information on alcohol use was included in the medical examination data for this cohort. There is a positive correlation between hypertension and alcohol consumption.[22] Other studies have shown that smokeless tobacco users consume more alcohol than nonusers.[19,20] Similarly, tobacco users tend to exhibit other risk behaviors more often than nonusers.[22] Considering the relative homogeneity of the study population—all of whom were construction industry employees—and the relative risk associated with smokeless tobacco use among subjects aged 35 through 54 years at entry into the study, as well as the relative risk associated with heavy alcohol consumption,[23] it is unlikely that differences in alcohol consumption (between smokeless tobacco users and nonusers) would explain the findings entirely. Moreover, the studies on the relationship between alcohol consumption and cardiovascular disease have been ambiguous.[24] There have been, however, no observations indicating a difference in alcohol consumption between smokeless tobacco users and cigarette smokers, and therefore the possible confounding effect of alcohol use can be suspected to be comparable for both kinds of tobacco use.

In a cross-sectional study of the health status of this cohort, systolic and diastolic blood pressure values were found to be significantly higher among smokeless tobacco users than among either nonusers or smokers.[10] This difference was not significant at younger ages but became obvious in subjects older than 45 years. Smokers had the lowest prevalence of hypertension, as has been found in many other epidemiological studies.[22,25,26] The significantly higher prevalence of hypertension found among smokeless tobacco users, compared with cigarette smokers, could reflect differences in the pharmacodynamic effects of nicotine in smokeless tobacco users and smokers. The use of smokeless tobacco yields high blood concentrations of nicotine and more prolonged and more perpetual levels than smoking. These high concentrations might be accompanied by longer periods of cutaneous vasoconstriction, systemic venoconstriction, and increased muscle blood flow in smokeless tobacco users,[27] and it could be speculated that such mechanisms are more operative for the development of hypertension in smokeless tobacco users than in smokers.

The construction industry demands a high level of physical performance. Persons with decreased capacity are not hired or leave the industry; therefore, construction workers are physically fitter than average citizens of the same age. Because the study sample was drawn from active workers and not from company lists, which might include retired workers, the healthy worker selection could be expected to be more obvious among the workers who belonged to the higher age group at entry into the study. This situation could explain the more pronounced excess risk of cardiovascular mortality found among the younger tobacco users than among the older tobacco users in relation to nonusers in the same age group.

Since smoking—and probably also the use of smokeless tobacco—affects the cardiovascular system in an unfavorable way and leads to poorer physical performance, the above-mentioned selection process may also be related to tobacco habit, probably resulting in an underestimation of the strength of the relationship between tobacco use and cardiovascular mortality.

The results of this study support the hypothesis that smokeless tobacco users face a higher risk than nonusers of dying of ischemic heart disease and cerebrovascular disorders. Smokers still seem to be at the highest risk, and quitting smoking

considerably reduces the risk of cardiovascular events. Even though nicotine is the main substance common to both smokers and smokeless tobacco users, it remains to be elucidated in clinical studies and by laboratory experiments whether the higher risk of cardiovascular disease in smokeless tobacco users is caused by nicotine or by some other exposure related to smokeless tobacco. Data on vasoconstriction, arrythmias, release of stress hormones, atherogenic effects, or hypertension caused by the exposure to nicotine in smokeless tobacco might elucidate this matter. The higher risk of dying of cardiovascular disease in smokers than in smokeless tobacco users supports the idea that there are substances in tobacco smoke other than nicotine that are hazardous. Although the risk of dying of cardiovascular disease is lower for smokeless tobacco users than for smokers, the excess risks are sufficiently impressive to call for action to be taken against the use of smokeless tobacco. □

Acknowledgments

This study was supported by grants from the Swedish Council for Social Research (B 91-0319:1) and the Swedish Heart and Lung Foundation.

Portions of this paper were presented at the Fourteenth Congress of the European Society of Cardiology, Barcelona, Spain, August 30–September 3, 1992.

References

1. *Reducing the Health Consequences of Smoking: 25 Years of Progress. A Report of the Surgeon General.* Rockville, Md: US Dept of Health and Human Services; 1989. DHHS publication CDC 89-8411.
2. Wilhelmsen L, Svärdsudd K, Korsan-Bengtsen K, Larsson B, Welin L, Tibblin G. Fibrinogen as a risk factor for stroke and myocardial infarction. *N Engl J Med.* 1984;311:501–505.
3. Pittilo RM, Clarke JM, Harris D, et al. Cigarette smoking and platelet adhesion. *Br J Haematol.* 1984;58:627–632.
4. Benowitz NL. Nicotine and cardiovascular disease. In: Adlkofer F, Thurau K, eds. *Effects of Nicotine on Biological Systems.* Basel, Switzerland: Birkhäuser Verlag; 1991;579–596. Advances in Pharmacological Sciences.
5. Lassila R, Seyberth HW, Haapanen A, Schweer H, Koskenvuo M, Laustiola KE. Vasoactive and atherogenic effects of cigarette smoking: a study of monozygotic twins discordant for smoking. *BMJ.* 1988; 297:955–957.
6. Pittilo RM. Cigarette smoking and endothelial injury: a review. *Adv Exp Med Biol.* 1990;273:61–78.
7. Ahmed SS, Moschos CB, Lyons MM, Oldewurtel HA, Coumbis RJ, Regan TJ. Cardiovascular effects of long-term cigarette smoking and nicotine administration. *Am J Cardiol.* 1976;37:33–40.
8. Gritz ER, Baer-Weiss V, Benowitz NL, Van Vunakis H, Jarvik ME. Plasma nicotine and cotinine concentrations in habitual smokeless tobacco users. *Clin Pharmacol Ther.* 1981;30:201–209.
9. Benowitz NL, Porchet H, Sheiner L, Jacob P. Nicotine absorption and cardiovascular effects with smokeless tobacco use: comparison with cigarettes and nicotine gum. *Clin Pharmacol Ther.* 1988;44: 23–28.
10. Bolinder G, Ahlborg B, Lindell J. Use of smokeless tobacco: blood pressure elevation and other health hazards found in a large-scale population survey. *J Intern Med.* 1992;232:327–334.
11. National Central Bureau of Statistics. Causes of death. In: *Official Statistics of Sweden 1985.* Stockholm, Sweden: Statistics Sweden.
12. Russell MA, Jarvis MJ, Devitt G, Feyerabend C. Nicotine intake by snuff users. *BMJ.* 1981;283:814–817.
13. Rothman KJ. *Modern Epidemiology.* Boston, Mass: Little, Brown & Co; 1986.
14. Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. *Biometrics.* 1985;41:55–68.
15. Huhtasaari F, Asplund K, Lundberg V, Stegmayr B, Wester PO. Tobacco and myocardial infarction: is snuff less dangerous than cigarettes? *BMJ.* 1992;305:1252–1256.
16. Wall S, Rosén M, Nyström L. The Swedish mortality pattern: a basis for health planning? *Int J Epidemiol.* 1985;14: 285–292.
17. de Faire U, Friberg L, Lorich U, Lundman T. A validation of cause-of-death certification in 1156 deaths. *Acta Med Scand.* 1976;200:223–228.
18. Poole C. Exceptions to the rule about non-differential misclassification. *Am J Epidemiol.* 1985;122:508.
19. Siegel D, Benowitz N, Ernster VL, Grady DG, Hauck WW. Smokeless tobacco, cardiovascular risk factors, and nicotine and cotinine levels in professional baseball players. *Am J Public Health.* 1992;82: 417–421.
20. Eliasson M, Lundblad D, Hägg E. Cardiovascular risk factors in young snuff-users and cigarette smokers. *J Intern Med.* 1991;230:17–22.
21. Tucker LA. Use of smokeless tobacco, cigarette smoking, and hypercholesterolemia. *Am J Public Health.* 1989;79:1048–1050.
22. Gyntelberg F, Meyer J. Relationship between blood pressure and physical fitness, smoking and alcohol consumption in Copenhagen males aged 40–59. *Acta Med Scand.* 1974;195:375–380.
23. Friedman LA, Kimball AW. Coronary heart disease mortality and alcohol consumption in Framingham. *Am J Epidem.* 1986;124:481–489.
24. Moore RD, Pearson TA. Moderate alcohol consumption and coronary artery disease: a review. *Medicine (Baltimore).* 1986;65:242–267.
25. Seltzer CC. Effect of smoking on blood pressure. *Am Heart J.* 1974;87:558–564.
26. Friedman GD, Klatsky AL, Siegelaub AB. Alcohol, tobacco and hypertension. *Hypertension.* 1982;III-4:143–150.
27. Benowitz NL. Clinical pharmacology of nicotine. *Annu Rev Med.* 1986;37:21–32.

Urinary Tract Infection during Pregnancy: Its Association with Maternal Morbidity and Perinatal Outcome

Laura A. Schieve, MS, Arden Handler, DrPH, Ronald Hershow, MD, Victoria Persky, MD, and Faith Davis, PhD

ABSTRACT

Objectives. **The effects of antepartum urinary tract infection on adverse maternal and perinatal outcomes were examined. Antepartum urinary tract infection has been previously implicated as a risk factor for numerous outcomes.**

Methods. **Crude and multivariable analyses were performed with a perinatal registry cohort of 25 746 mother/infant pairs.**

Results. **Elevated risks were observed for exposure to urinary tract infection and low birthweight, prematurity, preterm low birthweight, premature labor, hypertension/preeclampsia, maternal anemia, and amnionitis. Urinary tract infection was associated with perinatal death only among subjects 20 to 29 years of age.**

Conclusions. **These findings underscore the importance of antepartum urine screening to identify patients at risk for adverse outcomes. (*Am J Public Health.* 1994;84: 405–410)**

Introduction

Despite a plethora of research since Kass first noted an association between untreated urinary tract infection and perinatal mortality in 1962,[1] the extent to which antepartum urinary tract infection affects maternal and perinatal health remains ambiguous. At present, the only certain conclusion is that the risk of progression to pyelonephritis greatly increases during pregnancy, with rates as high as 30% reported among untreated women.[1–7]

Antepartum urinary tract infection has also been implicated as a risk factor for adverse perinatal outcomes—premature birth and/or low birthweight[1,3–5,7–14] and perinatal death[1,4,9]—in numerous observational studies. However, an equal number of negative findings have been reported.[2,6,15–28] Results of clinical trials examining the effect of antibiotic treatment on reducing the risk of low birthweight have been inconsistent as well.[1,3,4,7,9,10] In addition, associations have been documented between antepartum urinary tract infection and a variety of maternal complications of pregnancy, including hypertension/preeclampsia,[4,5,7,8,17] anemia,[8,9,15,22] amnionitis,[21] and endometritis.[29] The causal nature of these associations is questionable, because it is not always clear whether an episode of urinary tract infection preceded the particular outcome of interest, especially in regard to maternal hypertension and anemia. Although the definitions of these outcomes in many studies render it difficult to determine the proper sequence of events, a temporal relationship between urinary tract infection and hypertension was clearly established by Stuart et al.,[5] and a temporal relationship between urinary tract infection and anemia was established by Brumfitt.[9] These results support possible causal relationships. As with the perinatal outcomes, many investigators have also failed to demonstrate any effect with urinary tract infection and maternal outcomes.[2,6,11,16,20,24,27] Limited sample sizes and inadequate control of confounders may partially explain the discrepancy in findings to date.

This study used registry data to examine associations between antepartum urinary tract infection and adverse maternal and perinatal outcomes independent of other possible risk factors. The large sample size afforded by the registry also provided an opportunity to evaluate possible mechanisms through which urinary tract infection might exert an effect on pregnancy outcome.

Methods

We performed a retrospective cohort analysis with data from the University of Illinois Perinatal Network database, a registry of more than 150 000 obstetric events occurring from 1983 through 1989 in 14 hospitals in the Chicago area. Trained personnel abstracted more than 450 maternal, fetal, and neonatal variables from patients' medical records upon completion of pregnancy. We limited our analysis to 1988 and 1989 because of changes in abstracting format. We also limited the analysis to the four hospitals with the

The authors are with the School of Public Health, University of Illinois at Chicago.

Requests for reprints should be sent to Laura A. Schieve, MS, Program in Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, PO Box 6998, Chicago, IL 60680.

This paper was accepted June 2, 1993.

largest obstetric populations to allow a feasible reabstraction of a subset of records, in order to assess the exposure classification more thoroughly. All singleton live births and fetal deaths were included, with the exception of 115 subjects for whom information on urinary tract infection was not recorded (0.4% of the sample). In all, 25 746 women were included in this analysis; 1988 (7.7%) were considered positive for antepartum urinary tract infection and constituted the exposure group. The remainder constituted the nonexposed group.

Classification of urinary tract infection was based on (1) a positive urine culture during the antepartum period, as noted by a health care provider and recorded in the medical record, or (2) a diagnosis of urinary tract infection and/or pyelonephritis during the antepartum period recorded in the medical record. This allowed inclusion of cases of urinary tract infection diagnosed by a variety of methods, including culture, rapid-screening techniques, and clinical presentation. In most cases, it was not possible to discern the way in which a diagnosis was made from the information provided in the record. Our definition of urinary tract infection encompassed both asymptomatic and symptomatic infections because it included those individuals for whom a positive culture was recorded without a stated diagnosis of urinary tract infection. However, we were unable to determine the proportion of our exposed group that was symptomatic from the information provided.

The perinatal outcomes examined in this study include perinatal death (fetal death or neonatal death within the first 28 days of life, providing that the infant had not been discharged from the hospital), low birthweight (2500 g or lower) prematurity (less than 37 weeks gestation), preterm low birthweight (2500 g or lower *and* less than 37 weeks gestation), and small for gestational age (birthweight less than the 10th percentile of the birthweight distribution of the singleton deliveries in the database at each gestational age and infant sex).

Maternal outcomes evaluated include premature labor (onset of labor prior to 37 weeks of gestation, even if the episode was resolved and the pregnancy went to term), hypertension/preeclampsia (diagnosis of pregnancy-induced hypertension or preeclampsia specified by a health care provider), anemia (hematocrit less than 30%), amnionitis (diagnosis of amnionitis, chorioamnionitis, membranitis, or placentitis), and endometritis (diagnosis of endometritis or endomyometritis).[3] Because the exact timing of maternal complications was not recorded for this data set, it is unknown whether an episode of urinary tract infection actually preceded each of these outcomes. Therefore, this study is actually cross-sectional with respect to the maternal outcomes.

Because sexual intercourse is a risk factor for both urinary tract and genital tract infections and since genital tract infections may also be associated with adverse reproductive outcome, the detection of one or more genital tract pathogens during the antepartum period (syphilis, gonorrhea, genital herpes, condyloma, chlamydia, Beta Streptococcus vaginal infection, *Candida albicans*, *Trichomonas vaginalis*, *Gardnerella vaginalis*, unspecified vaginitis or cervicitis) was treated as a potential confounder in this analysis. Since ascertainment of antepartum urinary tract infection may have varied by hospital, odds ratios were also adjusted for hospital of delivery. Other variables evaluated as potential confounders include age (<20, 20 through 29, ≥30 years), race/ethnicity (White/Asian/other, Black, Hispanic), and prior pregnancy history. Pregnancy history was classified as follows: (1) primigravida, (2) at least one pregnancy ending in adverse outcome (congenital anomaly; fetal, neonatal, or postneonatal death; ectopic pregnancy or hydatid mole; spontaneous abortion; premature birth; or term birth with intrauterine growth retardation), (3) at least one pregnancy ending in induced abortion and no others ending in adverse outcomes or (4) all pregnancies ending in healthy outcomes (term births with no congenital anomalies or growth retardation).

Odds ratios (ORs) and 95% test-based confidence intervals (CIs) were computed for exposure to antepartum urinary tract infection and all perinatal and maternal outcomes. Mantel-Haenszel adjusted odds ratios were estimated from stratified analyses of all associations controlling for each of the potential confounders individually.[30] These summary estimates were evaluated with Breslow-Day tests of homogeneity.

Unconditional logistic regression was used to simultaneously adjust for age, race, prior pregnancy history, hospital of delivery (A, B, C, or D), and antepartum genital tract infection for each outcome. In addition, because univariate analysis revealed that the risk estimate for urinary tract infection and perinatal death was not homogeneous with respect to age, urinary tract infection and age interaction terms were entered into the model with perinatal death. Odds ratios were calculated from beta coefficients, and 95% confidence intervals were calculated from the respective standard errors.[31]

Because any association between urinary tract infection and adverse reproductive outcome may vary by the severity of infection, we repeated our analysis after subdividing the exposure group by progression to pyelonephritis. This analysis is presented only for the perinatal outcomes, however, since the number of subjects with both pyelonephritis and one of the maternal outcomes under study was very small and odds ratios were difficult to interpret. The relationship between the timing of infection and each outcome variable would also be of interest. Unfortunately, the data to address this issue were not available.

We also explored possible mechanisms by which urinary tract infection might affect perinatal outcome. Hypertension, anemia, and amnionitis were evaluated as possible intermediaries in the association between urinary tract infection and adverse perinatal outcome, because each of these conditions has been previously implicated as a consequence of antepartum urinary tract infection[4,5,7–9,15,17,21,22] and a risk factor for adverse perinatal outcome.[32–35] These analyses were limited to preterm low birthweight, the outcome with which urinary tract infection appeared most strongly associated, because any effect would be most likely with this association. We repeated both crude and multivariable analyses, excluding subjects with each of the postulated intermediary variables to assess whether the associations persisted in their absence.

Finally, in an effort to discern whether urinary tract infection per se is a risk factor for preterm low birthweight or whether the infectious process in general creates risk, the effect of urinary tract infection was compared with other infections commonly acquired by women of childbearing age, specifically genital and upper respiratory tract infections. The study population was divided into five mutually exclusive exposure groups: urinary tract infection only, genital tract infection only, upper respiratory tract

infection only, multiple infections (more than one of these three infections), and none of these infections. Odds ratios were calculated for each exposure group and preterm low birthweight, with the no infection group as the reference category.

Results

As shown in Table 1, the exposure groups were not homogeneous with respect to the potential confounding variables under study. Subjects classified in the urinary tract infection category were more likely than those in the nonexposed group to be non-White, to be younger, to have delivered at hospital D, and to have had an antepartum genital tract infection. They were less likely to have had a prior pregnancy that ended in a healthy outcome.

Women exposed to antepartum urinary tract infection were found to be at greater risk (unadjusted) of delivering infants with low birthweights, premature infants, preterm infants with low birthweights, and infants small for gestational age than those who were not exposed (Table 2). They were also more likely to experience premature labor, hypertension/preeclampsia, anemia, and amnionitis during pregnancy. Although the perinatal mortality and endometritis rates were higher among the exposure group, these differences did not reach statistical significance.

The risk of urinary tract infection on adverse perinatal outcomes was greatest among those with the most severe infection, pyelonephritis (Table 3). However, because only a small percentage of subjects (8.1%) progressed to pyelonephritis, the odds ratios after excluding these subjects from the exposure group were virtually the same as when exposure was defined as both pyelonephritis and nonpyelonephritis cases of urinary tract infection combined.

After single-factor adjustment for age, race, hospital of delivery, prior pregnancy history, and genital tract infections, all statistically significant associations in the crude analysis, with the exception of the urinary tract infection–small for gestational age association, remained elevated (data not shown). A significant association was also found between urinary tract infection and perinatal death when maternal age was between 20 and 29 years.

Multivariable adjustment reduced the risk estimates, but the associations between antepartum urinary tract infection and low birthweight, prematurity, preterm low birthweight, premature labor, hypertension, anemia, and amnionitis remained statistically significant (Table 4). As in the univariate analysis, the small for gestational age adjusted odds ratio was not significant, and the adjusted odds ratio for perinatal death was significant only for those 20 to 29 years old.

Neither the crude nor the adjusted odds ratio was modified when hypertension, anemia, or amnionitis cases were deleted from the urinary tract infection–preterm low birthweight analyses (adjusted OR excluding hypertension = 1.4, 95% CI = 1.2, 1.7; adjusted OR excluding anemia = 1.5, 95% CI = 1.2, 1.8; adjusted OR excluding amnionitis = 1.4, 95% CI = 1.1, 1.7). The risk of preterm low birthweight was virtually the same for women with an antepartum urinary tract infection only, an antepartum genital tract infection only, or both types of infection (Table 5). (Because of the unstable estimates from the multivariable analyses, only the unadjusted estimates are presented.) Women exposed to upper respiratory tract infection only were at no greater risk of delivering preterm low-birthweight infants than were those in the "infection-free" group.

TABLE 1—Selected Characteristics of Study Subjects, by Exposure to Antepartum Urinary Tract Infection

Characteristic	Exposed (n = 1988), %	Nonexposed (n = 23 758), %
Race*		
White/Asian/other	28.7	47.0
Black	30.1	17.8
Hispanic	41.1	35.3
Age*		
<20	22.6	13.9
20–29	57.5	54.5
30+	19.9	31.6
Hospital of delivery*		
A	17.5	36.6
B	28.9	25.5
C	15.3	15.9
D	38.3	22.1
Prior pregnancy history*		
Primigravida	30.7	29.4
Previous adverse outcome	27.4	24.0
Induced abortion (no adverse outcome)	11.9	10.0
All healthy outcomes	30.0	36.7
Genital tract infection*	21.2	11.3

*P < .001 by chi-square test for proportional differences for comparison between exposure groups.

Discussion

We found that women who acquire urinary tract infections during pregnancy are at increased risk of delivering low-birthweight, premature, and preterm low-birthweight infants. Our odds ratios of 1.4 for low birthweight, 1.3 for prematurity, and 1.5 for preterm low birthweight compare well with a recent meta-analysis of exposure to antepartum urinary tract infection that reported typical relative risks of 1.5 and 2.0 for associations with low birthweight and prematurity, respectively.[36] Antepartum urinary tract infection was not associated with delivering infants small for gestational age after multivariable adjustment for potential confounding variables, suggesting that urinary tract infection affects low birthweight through premature delivery rather than growth retardation. This is further illustrated by the relatively strong association between urinary tract infection and premature labor. Antepartum urinary tract infection was also found to be associated with increased maternal morbidity (hypertension, anemia, and amnionitis).

An association between urinary tract infection and perinatal death was found only among the 20- to 29-year-olds. One plausible explanation is that younger

TABLE 2—Unadjusted Odds Ratios for Exposure to Antepartum Urinary Tract Infection and Selected Perinatal and Maternal Outcomes

Outcome	Exposed	Nonexposed	Odds Ratio	95% Confidence Interval
Perinatal				
Perinatal death				
Yes	32	291	1.3	0.9, 1.9
No	1 956	23 467		
Birthweight, g				
≤2500	225	1 779	1.6	1.4, 1.8
>2500	1 759	21 942		
Gestational age, wk				
<37	293	2 546	1.4	1.3, 1.6
≥37	1 687	21 137		
Preterm low birthweight				
Yes	155	1 144	1.7	1.4, 2.0
No	1 821	22 505		
Small for gestational age				
Yes	258	2 529	1.3	1.1, 1.4
No	1 717	21 108		
Maternal				
Premature labor				
Yes	306	2 2'1	1.8	1.6, 2.0
No	1 679	21 547		
Hypertension				
Yes	122	973	1.5	1.3, 1.9
No	1 862	22 785		
Anemia				
Yes	123	716	2.1	1.8, 2.6
No	1 861	23 042		
Amnionitis				
Yes	54	413	1.6	1.2, 2.1
No	1 929	23 311		
Endometritis				
Yes	25	203	1.5	1.0, 2.2
No	1 957	23 525		

Note. Subjects missing information on an outcome variable are not included in that analysis. The maximum number missing is 134 (small for gestational age).

TABLE 3—Association between Antepartum Urinary Tract Infection and Selected Perinatal Outcomes, by Progression to Pyelonephritis

	Pyelonephritis		Nonpyelonephritis	
Outcome	Odds Ratio	95% Confidence Interval	Odds Radio	95% Confidence Interval
Perinatal death	2.6	1.1, 6.2	1.2	0.8, 1.8
Low birthweight	1.9	1.2, 2.9	1.5	1.3, 1.8
Prematurity	1.9	1.3, 2.9	1.4	1.2, 1.6
Preterm low birthweight	2.5	1.6, 4.0	1.6	1.3, 1.9
Small for gestational age	1.5	1.0, 2.3	1.2	1.1, 1.4

and older women are exposed to multiple competing risks for adverse outcomes—low socioeconomic status and related factors among the younger age group and biological factors among the older age group—which masks the impact of urinary tract infection.[37]

While several previous investigations have found comparable results, they have suffered from small sample sizes and were not adequately able to control for confounding. Of particular interest is whether the effects of urinary tract infection are independent of other infections with similar risk factors for acquisition, such as sexually transmitted diseases. We were able to adjust for antepartum sexually transmitted diseases and other genital tract pathogens in our analysis and did not find any change in the odds ratios. In addition, urinary tract infection and genital tract infection appear to have had similar effects on preterm low birthweight, as evidenced by the comparable odds ratios, when each type of infection was evaluated individually. However, we did not distinguish the genital infection category by specific pathogens. Our analysis also shows that the associations observed between urinary tract infection and both adverse perinatal outcome and maternal morbidity were not due to confounding by age, race, prior pregnancy history, or hospital of delivery.

A major weakness of this study is the method of exposure ascertainment. While most prior investigations have been conducted prospectively, with urine cultures on all women under investigation available to classify urinary tract infection, this study relied on a positive culture result or physician diagnosis of urinary tract infection and/or pyelonephritis in the medical record. This retrospective classification of urinary tract infection raised a number of concerns about the validity and reliability of this variable. Although it was impossible for us to assess validity directly, we reabstracted a random sample of charts from our study population to evaluate the reliability of the original abstraction process and to provide some sense of potential limitations of the urinary tract infection variable beyond abstraction error. In all, 217 charts were reabstracted, 112 from the exposed group (5%) and 105 from the nonexposed group (0.4%).

Agreement between this second abstraction and the original classification of urinary tract infection in the database was 96.2% for the nonexposed group but only 69.6% for the exposure group. (That is, 30.4% of our exposure group was not reclassified as having a urinary tract infection, during the second abstraction.) Further evaluation of these apparently misclassified subjects revealed that the incidence of adverse outcomes was lower in this group than for those classified correctly as having a

TABLE 4—Adjusted Odds Ratios for Exposure to Antepartum Urinary Tract Infection and Selected Perinatal and Maternal Outcomes

Outcome	Odds Ratio	95% Confidence Interval
Perinatal		
Perinatal death		
Age < 20 d	0.3	0.1, 1.3
Age 20–29 d	1.8	1.1, 2.8
Age 30+ d	1.1	0.5, 2.3
Low birthweight	1.4	1.2, 1.6
Prematurity	1.3	1.1, 1.4
Preterm low birthweight	1.5	1.2, 1.7
Small for gestational age	1.1	0.9, 1.3
Maternal		
Premature labor	1.6	1.4, 1.8
Hypertension	1.4	1.2, 1.7
Anemia	1.6	1.3, 2.0
Amnionitis	1.4	1.1, 1.9
Endometritis	1.0	0.6, 1.5

Note. Estimates were adjusted for race, age, prior pregnancy history, hospital of delivery, and genital tract pathogens by means of unconditional logistic regression. Subjects missing information on outcome or confounding variables are not included in that analysis. The maximum number missing is 357 (small for gestational age). Stratum-specific estimates were adjusted for race, prior pregnancy history, hospital of delivery, and genital tract pathogens.

TABLE 5—Unadjusted Odds Ratios for Exposure to Antepartum Urinary Tract Infection (UTI), Genital Tract Infection, Upper Respiratory Tract Infection (URI), and Combinations of These Infections for Preterm Low Birthweight

Exposure	n	Odds Ratio	95% Confidence Interval
Negative for all 3 infections	19 927	1.0	. . .
UTI only	1 464	1.8	1.4, 2.1
Genital infection only	2 475	1.5	1.3, 1.8
URI only	1 053	0.8	0.6, 1.2
Multiple infections	706	1.6	1.2, 2.2
UTI and URI	94	1.4	0.6, 3.3
UTI and genital infection	394	1.9	1.3, 2.7
URI and genital infection	194	1.2	0.6, 2.3
All 3 infections	24	0.9	0.1, 6.1

urinary tract infection. Therefore, including these subjects in the exposure group biased our results toward the null hypothesis, and the true associations are likely to be stronger than our estimates indicate.

In addition to abstraction quality, the classification of urinary tract infection used in this study is subject to a number of other potential shortcomings. First, if a woman did not receive prenatal care, an episode of urinary tract infection may have occurred during pregnancy that went unrecorded. In addition, even if a woman did receive prenatal care, her prenatal care record may not have been incorporated into her hospital of delivery medical record, potentially resulting in a missed case of urinary tract infection. Our reabstraction revealed that 13.4% of our study population did not receive prenatal care and that the prenatal care record was unavailable for an additional 4.8%.

In addition to receipt of prenatal care, we were also concerned about the prevalence of laboratory screening for urinary tract infection. Although most prenatal care regimes include an antepartum urinalysis, not all routinely incorporate urine cultures. We found that antepartum urine cultures were recorded for only 33.2% of our reabstracted sample. This is most likely an underestimate, because it is feasible that urine cultures performed during the antepartum period were not always recorded in the medical record. However, it is still cause for concern, since urinary tract infection may be asymptomatic and urinalysis alone may be insensitive. Although these data suggest that cases of asymptomatic urinary tract infection may have gone undetected in this population, any consequent bias would again be toward the null hypothesis because comparison groups would be more homogeneous on exposure.

Given the potential errors in classification, it is encouraging to note that the prevalence of urinary tract infection at each hospital under study fell within the 2% to 10% range seen in other investigations.[1,3,5,6,12,15,26] This rate was actually exceeded at one hospital that serves a large number of indigent women, an observation also consistent with current literature.

Although the registry does not contain information on antibiotic treatment for urinary tract infection, we found, during our reabstraction, that antibiotics were prescribed for 84.6% of those correctly classified as having such an infection, indicating that the exposure group involves mainly treated cases.

Assuming that the association between urinary tract infection and prematurity is real, the mechanism by which the former might exert such an effect is not well established. Several investigators have demonstrated a high incidence (up to 50%) of pyelographic abnormalities, indicative of chronic pyelonephritis, in bacteriuric women.[4,38] It is plausible, then, that the effect of urinary tract infection on premature birth could be indirectly mediated by occult renal abnormalities that result in antepartum maternal complications such as hypertension/preeclampsia or anemia. Our data do not support this hypothesis, because no change in the odds ratio was observed for the association between urinary tract infection and preterm low birthweight when women with either hypertension or anemia were excluded from analysis. We cannot eliminate other possible effects of renal abnormalities, however.

It is also conceivable that urinary tract infection affects premature labor directly, through the development of amnionitis. It has been previously suggested that bacterial infection of the amniotic fluid is a risk factor for premature delivery.[34,35] One hypothesis contends that bacterial enzymes such as collagenase may weaken the fetal membranes and predispose them to rupture, which subsequently triggers the onset of labor.[39] It has also been postulated that bacterial products such as phospholipase A and C or endotoxins may stimulate prostaglandin biosynthesis by the fetal membranes, which then initiates labor.[40] Alternatively, these bacterial products might stimulate the patient's immune system, causing the release of monokines (particularly Platelet Activating Factor, Interleukin-1, and Tumor Necrosis Factor) that then trigger prostaglandin production.[40] It has also been demonstrated that bacteriuric women may subsequently develop an amniotic fluid

infection with the same organisms commonly isolated from urine specimens, namely coliform bacteria such as *Escherichia coli*.[21] We found no change in the odds ratio after deleting cases of amnionitis, although subclinical amniotic infection has been repeatedly documented[41-43] and may have been a latent causative factor in this population.

Future studies are needed to address the interrelationship between urinary tract infection, amnionitis, and premature labor. Ideally, these associations should be explored in a prospective study with well-defined criteria for both exposure and outcome variables.

Although the mechanism remains unclear, these data, along with prior investigations, indicate that antepartum urinary tract infection indeed affects both maternal and perinatal health, even when treated. Early prenatal screening of women for urinary tract infection and treatment of women found to be bacteriuric are already indicated to prevent progression to pyelonephritis. The associations documented here provide further compelling support for screening to identify patients at risk for adverse pregnancy outcomes. □

Acknowledgments

We are grateful to the University of Illinois Perinatal Management Group for allowing us access to the database. In addition, we would like to thank Deborah Rosenberg and Cynthia Ferre for their assistance with computer programming and Susan Ataman, Susan Johnson, and Edna Brooks for their help with data collection.

References

1. Kass EH. Pyelonephritis and bacteriuria: a major problem in preventive medicine. *Ann Intern Med.* 1962;56:46–53.
2. Bryant RE, Windom RE, Vineyard JP, Sanford JP. Asymptomatic bacteriuria in pregnancy and its association with prematurity. *J Lab Clin Med.* 1964;63:224–231.
3. LeBlanc AL, McGanity WJ. The impact of bacteriuria in pregnancy—a survey of 1300 pregnant patients. *Biol Med.* 1964;22: 336–347.
4. Kincaid-Smith P, Bullen M. Bacteriuria in pregnancy. *Lancet.* 1965;1:395–399.
5. Stuart KL, Cummins GTM, Chin WA. Bacteriuria, prematurity, and the hypertensive disorders of pregnancy. *Br Med J.* 1965;1:554–556.
6. Little PJ. The incidence of urinary infection in 5000 pregnant women. *Lancet.* 1966;4:925–928.
7. Wren BG. Subclinical urinary infection in pregnancy. *Med J Aust.* 1969;1:1220–1226.
8. Naeye RL. Urinary tract infections and the outcome of pregnancy. *Adv Nephrol.* 1986;15:95–102.
9. Brumfitt W. The effects of bacteriuria in pregnancy on maternal and fetal health. *Kidney Int.* 1975;8:S113–S119.
10. Savage WE, Hajj SN, Kass EH. Demographic and prognostic characteristics of bacteriuria in pregnancy. *Medicine.* 1967; 46:385–407.
11. Condie AP, Williams JD, Reeves DS, Brumfitt W. Complications of bacteriuria in pregnancy. In: O'Grady F, Brumfitt W, eds. *Urinary Tract Infection.* London, England: Oxford University Press; 1968: 148–159.
12. Gruneberg RN, Leigh DA, Brumfitt W. Relationship of bacteriuria in pregnancy to acute pyelonephritis, prematurity, and fetal mortality. *Lancet.* 1969;2:1–3.
13. McGrady GA, Daling JR, Peterson DR. Maternal urinary tract infection and adverse fetal outcomes. *Am J Epidemiol.* 1985;121:377–381.
14. Henderson M, Reinke WA. The relationship between bacteriuria and prematurity. In: Kass EH, ed. *Progress in Pyelonephritis.* Philadelphia, Pa: FA Davis; 1965: 27–32.
15. Layton R. Infection of the urinary tract in pregnancy: an investigation of a new routine in antenatal care. *J Obstet Gynaecol Br Commonwealth.* 1964;71:927–933.
16. Sleigh JD, Robertson JG, Isdale MH. Asymptomatic bacteriuria in pregnancy. *J Obstet Gynaecol Br Commonwealth.* 1964; 71:74–81.
17. Norden CW, Kilpatrick WH. Bacteriuria of pregnancy. In: Kass EH, ed. *Progress in Pyelonephritis.* Philadelphia, Pa: FA Davis; 1965:64–72.
18. Whalley PJ. Bacteriuria in pregnancy. In: Kass EH, ed. *Progress in Pyelonephritis.* Philadelphia, Pa: FA Davis; 1965:50–57.
19. Wilson MG, Hewitt WL, Monzon OT. Effect of bacteriuria on the fetus. *N Engl J Med.* 1966;274:1115–1118.
20. Dixon HG, Brant HA. The significance of bacteriuria in pregnancy. *Lancet.* 1967;1: 19–20.
21. Patrick MJ. Influence of maternal renal infection on the foetus and infant. *Arch Dis Child.* 1967;42:208–213.
22. Robertson JG, Livingstone JRB, Isdale MH. The management and complications of asymptomatic bacteriuria in pregnancy. *J Obstet Gynaecol Br Commonwealth.* 1968;75:59–65.
23. Elder HA, Santamarina BAG, Smith S, Kass EH. The natural history of asymptomatic bacteriuria during pregnancy: the effect of tetracycline on the clinical course and the outcome of pregnancy. *Am J Obstet Gynecol.* 1971;111:441–462.
24. Schamadan WE. Bacteriuria during pregnancy. *Am J Obstet Gynecol.* 1964;89:10–15.
25. Carroll R, MacDonald D, Stanley JC. Bacteriuria in pregnancy. *Obstet Gynecol.* 1968;32:525–527.
26. Bobeck S, Schersten B. Detection and diagnosis of bacteriuria in pregnancy. *Practitioner.* 1974;212:257–262.
27. Gilstrap LC, Leveno KJ, Cunningham FG, Whalley PJ, Roark ML. Renal infection and pregnancy outcome. *Am J Obstet Gynecol.* 1981;141:709–716.
28. Reddy J, Cambell A. Bacteriuria in pregnancy. *Aust NZ J Obstet Gynaecol.* 1985;25:176–178.
29. Monif GRG. Intrapartum bacteriuria and postpartum endometritis. *Obstet Gynecol.* 1991;78:245–248.
30. Rothman KJ. *Modern Epidemiology.* Boston, Mass: Little, Brown; 1986:177–236.
31. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: Wiley; 1989:38–81.
32. Kramer MS, McLean FH, Eason EL, Usher RH. Maternal nutrition and spontaneous preterm birth. *Am J Epidemiol.* 1992;136:574–583.
33. Lieberman E, Ryan KJ, Monson RR, Schoenbaum SC. Association of maternal hematocrit with premature labor. *Am J Obstet Gynecol.* 1988;159:107–114.
34. Chellam VG, Rushton DI. Chorioamnionitis and funiculitis in the placentas of 200 births weighing less than 2.5 kg. *Br J Obstet Gynaecol.* 1985;92:808–814.
35. Guzick DS, Winn K. The association of chorioamnionitis with preterm delivery. *Obstet Gynecol.* 1985;65:11–16.
36. Romero R, Oyarzun E, Mazor M, Sirtori M, Hobbins JC, Bracken M. Meta-analysis of the relationship between asymptomatic bacteriuria and preterm delivery/low birth weight. *Obstet Gynecol.* 1989;73:576–582.
37. Institute of Medicine. *Preventing Low Birthweight.* Washington, DC: National Academy Press; 1985:46–93.
38. Diokno AC, Compton A, Seski J, Vinson R. Urologic evaluation of urinary tract infection in pregnancy. *J Reprod Med.* 1986;31:23–26.
39. Cox SM. Infection-induced preterm labor. In: Gilstrap LC, Faro S, eds. *Infections in Pregnancy.* New York, NY: Alan R Liss Inc; 1990:247–253.
40. Romero R, Mazor M. Infection and preterm labor. *Clin Obstet Gynecol.* 1988; 31:553–584.
41. Miller JM, Hill GB, Welt SI, Pupkin MJ. Bacterial colonization of amniotic fluid in the presence of ruptured membranes. *Obstet Gynecol.* 1980;137:451–458.
42. Bobitt JR, Havslip CC, Damato JD. Amniotic fluid infection as determined by transabdominal amniocentesis in patients with intact membranes in premature labor. *Am J Obstet Gynecol.* 1981;140:947–952.
43. Leigh J, Garite TJ. Amniocentesis and the management of premature labor. *Obstet Gynecol.* 1986;67:500–506.

Maternal Attitude toward Pregnancy and the Risk of Neonatal Death

Muhammad N. Bustan, MD, DrPH, and Ann L. Coker, PhD

ABSTRACT

Objectives. **Reduced options for fertility control over the past decade have increased the rates of unwanted pregnancy. We evaluated whether a woman's negative attitude toward her pregnancy increased the risk of perinatal mortality, in a large, prospective cohort study.**

Methods. **The association between attitude toward the pregnancy and perinatal mortality was evaluated in a longitudinal cohort study of 8823 married, pregnant patients enrolled from 1959 to 1966 in the Child Health and Development Studies.**

Results. **Women who reported during the first trimester of prenatal care that the pregnancy was unwanted were more than two times more likely to deliver infants who died within the first 28 days of life than were women reporting accepted pregnancies. A positive attitude toward pregnancy was not associated with fetal death or postneonatal death.**

Conclusions. **These data, collected when induced abortions were illegal, may have important implications for the 1990s. If maternal attitude toward the pregnancy is associated with neonatal mortality and abortion laws change such that access is restricted, infant mortality may increase because a greater proportion of births will be unwanted. (*Am J Public Health.* 1994; 84:411–414)**

Introduction

In 1973, the year of the *Roe v Wade* Supreme Court decision (410 US 113) that made abortion legal in the United States, 14.3% of pregnancies were reported by respondents to the National Survey of Family Growth as unwanted at conception.[1] Almost 10 years later, in 1982, the percentage of pregnancies reported as unwanted at conception fell to almost one half the 1973 percentage (7.7%). However, during the late 1980s, when abortion remained legal but politically and legally challenged, the percentage of women reporting pregnancies as unwanted increased to 10.3%.[1] Given this trend toward increasing numbers of unwanted pregnancies, particularly if abortion becomes inaccessible in the United States, further attention must be given to the consequences of unwanted childbearing.

Being unwanted puts children at increased risk of a range of adverse health outcomes, including child abuse[2] and delayed cognitive and social-emotional development.[3–5] Women with unwanted pregnancies are less likely to seek early and adequate prenatal care[6–8] and may thus be at increased risk of delivering infants with adverse pregnancy outcomes. In a Finnish study, women reporting infants unwanted at birth were more likely to have infants of lower birthweights and higher rates of infant mortality.[9] Kafatos and Pantelakis reported that being illegitimate or unwanted was a risk factor for perinatal morbidity and mortality.[10]

No US-based study of pregnancy attitude and perinatal mortality has assessed attitude early in the pregnancy and followed women to delivery to ascertain pregnancy outcomes. This analysis of the Child Health and Developmental Studies cohort (recruited from 1959 to 1965) assessed attitude toward the pregnancy at the first prenatal care visit and followed these Kaiser Health Plan patients to delivery. The risk of fetal, neonatal, and postneonatal mortality associated with being unwanted was estimated while controlling for important confounders, including level of prenatal care.

Methods

The Child Health and Development Studies involved a prospective cohort of 20 754 pregnancies to women who were members of the Kaiser Health Plan in the East Bay Area of San Francisco between June 1959 and September 1966.[11] Women attending Kaiser clinics for confirmation of pregnancy were asked to participate. Pregnant women were then followed throughout their pregnancy while receiving prenatal care in Kaiser clinics. Those who changed care providers during the pregnancy or delivered at facilities other than Kaiser's were dropped from the follow-up. Subsets of children were followed for up to 15 years.

During the pregnancy follow-up periods, information from several sources was collected. Women were to be interviewed early in their pregnancies to

The authors are with the Department of Epidemiology and Biostatistics, School of Public Health, University of South Carolina, Columbia. Muhammad N. Bustan is also with the School of Public Health, Hasanuddin University, Ujungpandang, Indonesia.

Requests for reprints should be sent to Ann L. Coker, PhD, Department of Epidemiology and Biostatistics, School of Public Health, University of South Carolina, Columbia, SC 29208.

This paper was accepted June 29, 1993.

obtain information on their medical history; behavioral risk factors, including tobacco and alcohol intake; and sociodemographic factors. Also, the women and their husbands were asked about their attitude toward the pregnancy. Existing medical records were computerized and linked with the mother's personal interview, including the prenatal care record, and the mother's and infant's hospital records.

For the purposes of this analysis, women who were unmarried, who were interviewed after delivery, or who did not answer the question on attitude toward the pregnancy were excluded. In addition, to address the issue of nonindependence of pregnancy outcomes with the same mother, only one randomly selected pregnancy per woman was included in this analysis, leaving a total of 8823 eligible married women 20 to 50 years old.

Attitude toward the pregnancy was classified on the basis of the woman's response to a question asked during the first trimester of pregnancy about her perception and that of her husband in regard to their pregnancy. If the woman indicated that either she or her husband was unhappy, resentful, or upset about the pregnancy or did not want the pregnancy or that the pregnancy was mistimed, the pregnancy was categorized as unwanted. If they indicated that they were happy about the pregnancy or stated that they wanted, accepted, or planned the pregnancy, the pregnancy was categorized as accepted. Women who reported that they or their husband had contradictory feelings about the pregnancy were not included in either group for this analysis. Changes in attitudes during pregnancy were not considered.

Information regarding fetal, neonatal, and postneonatal death was obtained from medical records and autopsy reports. A fetal death was defined as fetal loss prior to delivery and included stillbirths as well as recognized early pregnancy loss. A neonatal death was defined as a live-born baby who expired within 28 days of birth, while a postneonatal death was defined to include live-born infants dying after the 28th day of life but prior to 1 year of age.[12] The specific cause of death was not noted.

Information on the following risk factors was available from data collected for this cohort of women: maternal and paternal age, race, educational attainment, religion, family income, parity, interpregnancy interval, maternal and paternal smoking and drinking exposure, infant birthweight, and the presence of a congenital anomaly in the infant. These data were used to identify factors that may confound the relationship between being unwanted and infant mortality (fetal, neonatal, and postneonatal mortality). Factors associated with attitude toward the pregnancy were first described by comparing the proportion of those reporting an unwanted pregnancy with the selected risk factors and the proportion of those reporting an accepted pregnancy. Similarly, risk factors for neonatal mortality were assessed by comparing the proportion of those experiencing neonatal mortality with the selected risk factors and the proportion among live-born and surviving infants. Factors associated with both attitude toward the pregnancy and neonatal mortality (confounders)[13] were included in logistic regression models estimating the risk of infant death associated with being unwanted.

Results

Of the 8823 married, pregnant Kaiser patients enrolled in the study and included in this analysis, 1274 (14.4%) reported that the pregnancy was unwanted when asked during the first trimester. Several factors were associated with reporting the pregnancy as unwanted vs wanted. In comparison with women reporting the pregnancy as accepted, women reporting that the pregnancy was unwanted were 2.9 times more likely (95% confidence interval [CI] = 2.2, 3.7) to be 40 years of age or older, twice as likely (95% CI = 1.7, 2.3) to have husbands with less than a high school education, and 3.7 times more likely (95% CI = 3.1, 4.4) to have had four or more children. Women reporting an unwanted pregnancy were 40% more likely (odds ratio [OR] = 1.4, 95% CI = 1.3, 1.6) to be non-White and 20% more likely (OR = 1.2, 95% CI = 1.0, 1.4) to be smokers. Reporting an unwanted pregnancy was not associated with the infant having had any congenital anomaly (OR = 1.0, 95% CI = 0.7, 1.4), with having a short interpregnancy interval among parous women (OR = 0.9, 95% CI = 0.8, 1.1), or with low infant birthweight (OR = 1.1, 95% CI = 0.9, 1.4). In addition, maternal education, paternal age, and paternal race were not independently associated with attitude toward the pregnancy (when controlling for paternal education, maternal age, and race).

Although information was collected on other potential confounders, sufficiently large numbers of missing values precluded the use of family income (10% missing), religion (30% missing), and maternal and paternal alcohol consumption (15% missing).

Of the cohort of 8823 live births, 83 infants (0.9%) died within the first 28 days of life (neonatal deaths) and 57 (0.6%) died after the first 28 days but prior to 1 year of age (postneonatal deaths). Three hundred thirty-one fetal deaths (3.8%) were observed in this cohort of pregnancies. Risk of neonatal death increased with decreasing educational level of the husband (OR = 1.9 for less than a high school education, 95% CI = 0.9, 3.7) and with increasing parity (OR = 2.4 for parity of four or more, 95% CI = 1.2, 4.5). As expected, low birthweight (< 2500 g) was strongly associated with neonatal death (OR = 39.6, 95% CI = 29.6, 52.9), as was having any congenital anomaly (OR = 10.7, 95% CI = 7.2, 15.9). Women 40 years of age or older appeared to be at a slightly reduced risk of having a child who died prior to 28 days, but the confidence interval for this estimate was wide (OR = 0.4, 95% CI = 0.02, 2.58). Maternal race, interpregnancy interval, and smoking during pregnancy were not consistently associated with neonatal death. Also, maternal education, paternal age, and paternal race were not independently associated with neonatal death.

Table 1 presents the number and percentage of unwanted and accepted pregnancies by the following pregnancy outcome categories: fetal death, neonatal death, postneonatal death, and live-born infants surviving beyond 1 year of age (n = 8740). Crude and adjusted relative risk estimates for pregnancy attitude and the perinatal mortality outcomes obtained from multiple logistic regression are also presented in Table 1. Only neonatal death was significantly associated with being unwanted. Women reporting, during the first trimester of prenatal care, that the pregnancy was not wanted were 2.4 times more likely (95% CI = 1.5, 4.0) to deliver infants who died within the first 28 days of life. The crude relative risk estimate for this relationship was 2.7 (95% CI = 1.8, 4.2). Women reporting their pregnancy as unwanted were no more likely to have the pregnancy end in a

fetal death (adjusted OR = 1.1, 95% CI = 0.9, 1.5) than were women accepting their pregnancies. Postneonatal death was not associated with attitude toward the pregnancy; the adjusted relative risk estimate was 0.9 (95% CI = 0.5, 1.9).

TABLE 1—Relative Risks for the Relationship between Maternal Attitude toward the Pregnancy and Perinatal and Infant Death

	Unwanted Pregnancy		Accepted Pregnancy		Crude RR (95% CI)	Adjusted RR[a] (95% CI)
	No.	%	No.	%		
Fetal death	62	4.6	269	3.4	1.4 (1.0, 1.8)	1.1 (0.9, 1.5)
Death within 1–27 d	25	1.9	57	0.7	2.7 (1.8, 4.3)	2.4 (1.5, 4.0)
Death after 28 d	7	0.5	50	0.6	0.8 (0.4, 1.9)	0.9 (0.5, 1.9)
Live birth	1248	92.9	7492	95.2	1.0 Reference	1.0 Reference

Note. RR = relative risk; CI = confidence interval.
[a]Adjusted for gravida's age, race, and parity and husband's education.

Discussion

Parents (either the mother or father) perceiving their pregnancy as unwanted had a greater than twofold increased risk of delivering a child who died within the first 28 days of life. Attitude toward the pregnancy was not associated with an increased risk of fetal death or with an increased risk of postneonatal mortality.

Few studies have directly addressed the relationship between pregnancy attitude and infant mortality,[9,10] and none have assessed attitude early in the pregnancy and followed women to determine subsequent pregnancy outcomes. Kafatos and Pantelakis[10] did find a significant association between illegitimacy and perinatal mortality, while Myrhman[9] found that being unwanted increased the risk of low birthweight but not perinatal mortality.

Women enrolled in the Child Health and Development Studies were a select group with unique characteristics that placed them at a reduced risk of adverse pregnancy outcomes. Since all women received early prenatal care, lack of access to such care cannot explain the observed relationship between attitude toward the pregnancy and neonatal mortality. All women in this cohort were married and were either employed or married to men whose employment guaranteed the Kaiser Health Plan. Therefore, the proportion of women who reported unwanted pregnancies in this cohort was probably lower than in a general population of pregnant women. In comparison with a population having less access to prenatal care, this cohort was at reduced risk of adverse pregnancy outcomes, including neonatal mortality.

In the 1990s, relative to the 1960s when these data were collected, significant improvements in survival rates for high-risk, primarily low-birthweight infants have been achieved. In the data reported here, being unwanted was not associated with low birthweight (OR = 1.1); thus, the observed association between attitude toward the pregnancy and infant mortality should still be applicable for pregnancies in the 1990s.

Our finding of no association between attitude toward the pregnancy and postneonatal mortality is somewhat puzzling. However, because the cause of death was not coded, we cannot, for example, look at the relationship between pregnancy attitude and intentional injury, nor can we correlate time of death with its cause by pregnancy attitude. Since this is one of the first studies to report an increased risk of neonatal death associated with being unwanted, additional cohort studies should be conducted in an attempt to replicate these findings.

The definition of an unwanted pregnancy used in this analysis differed from those used in other surveys[14] and may be less biased because women were asked about their feelings regarding the pregnancy at least 6 months prior to delivery. In addition, the question addressed both the woman's and her husband's perceptions of the pregnancy (illegitimate births were not included). Finally, women were allowed a range of responses to this question characterizing their feeling about the pregnancy; pregnancy attitude was not asked as part of the question but was included as a response option.

During the time that women were enrolled in this study (1959 to 1966), abortion was illegal and contraception was not widely available. In light of current political challenges to abortion and family planning, these data may have a particularly important message. If access to abortion and family planning is limited or eliminated, the proportion of women delivering unwanted infants will increase dramatically. If pregnancy attitude is linked to neonatal death and the proportion of unwanted infants increases, then rates of infant mortality may also increase. Since the women included in the analysis were primarily middle income and received early prenatal care and were thus at reduced risk of adverse pregnancy outcomes, the impact of being unwanted on infant mortality may be underestimated. Thus, if these results are applied to women without prenatal care, the magnitude of the effect of attitude toward the pregnancy on infant mortality may be greater.

Pragmatically, reducing the proportion of infants born unwanted is a public health issue that we know how to and have had success in addressing. Public health interventions (family planning and abortion services) do exist to increase the proportion of wanted births and, thus, improve a range of infant health outcomes. □

References

1. Williams LB. Determinants of unintended childbearing among ever-married women in the United States: 1973–1988. *Fam Plann Perspect.* 1991;23:212–222.
2. Zuravin SJ. Unplanned childbearing and family size: their relationship to child neglect and abuse. *Fam Plann Perspect.* 1991;23:155–161.
3. Baldwin W, Cain VS. The children of teenage parents. *Fam Plann Perspect.* 1980;12:34–40.
4. Matejcek Z. The Prague cohort through age nine. In: David HP, Dytrych Z, Matejcek Z, Schuller V, eds. *Born Unwanted.* New York, NY: Springer; 1988: 53–86.
5. David HP. Additional studies from Sweden. In: David HP, Dytrych Z, Matejcek Z, Schuller V, eds. *Born Unwanted.* New York, NY: Springer; 1988:46–52.
6. Lia-Hoagberg B, Rode P, Skovholt C, et al. Barriers and motivators to prenatal care among low-income women. *Soc Sci Med.* 1990;30:487–495.
7. Sable MR, Stockbauer JW, Schramm WF, Land GH. Differentiating the barriers to adequate prenatal care in Missouri, 1987–1988. *Public Health Rep.* 1990;105: 549–555.
8. Weller RH, Eberstein IW, Bailey M.

Pregnancy wantedness and maternal behavior during pregnancy. *Demography.* 1987;24:407–412.
9. Myhrman A. The Northern Finland cohort, 1966–82: a follow-up study of children unwanted at birth. In: David HP, Dytrych Z, Matejcek Z, Schuller V, eds. *Born Unwanted.* New York, NY: Springer; 1988:103–110.
10. Kafatos AG, Pantelakis SN. Factors related to perinatal morbidity and mortality. *Paediatrician.* 1982;11:27–44.
11. Yerushalmy J. Biostatistical methods in investigations of child health. *Am J Dis Child.* 1967;114:470–476.
12. Bakketeig LA, Hoffman HJ, Titmuss Oakley AR. Perinatal mortality. In: Bracken MB, ed. *Perinatal Epidemiology.* New York, NY: Oxford University Press; 1984:99–151.
13. Mausner JS, Bahn AK, Kramer S. *Epidemiology: An Introductory Text.* Philadelphia, Pa: WB Saunders; 1985.
14. Klerman LV, Jekel JF. Unwanted pregnancy. In: Bracken MB, ed. *Perinatal Epidemiology.* New York, NY: Oxford University Press; 1984:283–299.

Women's Health Conference to Be Held in May

The American Psychological Association will hold a conference entitled "Psychosocial and Behavioral Factors in Women's Health: Creating an Agenda for the 21st Century" May 12 through May 14, 1994, at the Capitol Hill Hyatt Regency Hotel, Washington, DC. There will be a preconference awards reception on Capitol Hill on the evening of May 11, 1994, to honor legislators and other key figures in women's health. The goal of the conference is to highlight the importance of psychosocial and behavioral factors in women's health research and the implications for treatment, prevention, and health policy.

Major foci will include theoretical models/frameworks for conceptualizing women's health; issues in research methodology, measurement, and evaluation; new research on psychosocial and behavioral factors in women's health; implications of psychological factors in treatment, health policy, and interventions; and special issues of underserved populations (e.g., ethnic minorities, the poor, women with disabilities, lesbians, and older women). Content areas of interest will include sociocultural influences on health; behavioral and psychosocial risk factors; behavioral and psychosocial factors in health promotion; and coping, resilience, health, and illness.

Registration fees for professionals are $225 (preregistration [received by April 8, 1994]) or $275 (late and on-site registration). Fees for students are $95 (preregistration [received by April 8, 1994]) or $125 (late and on-site registration).

For more information, please contact Wanda Robinson, Women's Health Conference, American Psychological Association, Women's Programs Office, 750 First St, NE, Washington, DC 20002-4242; tel (202) 336-6120; fax (202) 336-6117.

A 6-Year Follow-Up of Behavior and Activity Disorders in the Taiwan Yu-cheng Children

Yung-Cheng Joseph Chen, MD, MPH, Mei-Lin M. Yu, MD, DrPH, Walter J. Rogan, MD, MPH, Beth C. Gladen, PhD, and Chen-Chin Hsu, MD

ABSTRACT

Objectives. **The relationship of behavior and activity levels to the interval between outbreak and year of birth and to age of the children is explored in Taiwanese children exposed in utero to heat-degraded polychlorinated biphenyls (PCBs)—the Yu-cheng children. Additionally, the relationship of the scores to chemical, physical, and cognitive findings is described.**

Methods. **With Rutter's Child Behavior Scale A and a modified Werry-Weiss-Peters Activity Scale, 118 Yu-cheng children and matched controls were followed biannually from 1985 to 1991.**

Results. **At each year, the Yu-cheng children scored 7% to 43% worse (mean = 23%) than control children on the Rutter scale. At any fixed age, the Yu-cheng children scored 11% to 63% (mean = 28%) worse. The effect for children born later did not differ from that for those born earlier; neither was there any improvement as the children aged. A similar but weaker picture was seen for the activity score. These behavioral findings were not related to physical or cognitive findings or to serum PCB levels.**

Conclusions. **In utero exposure to heat-degraded PCBs appears to cause mildly disordered behavior and increased activity level; the effect persists over time and is similar in children born up to 6 years after the mothers were exposed. (*Am J Public Health.* 1994;84: 415–421)**

Introduction

Polychlorinated biphenyls (PCBs), a group of compounds with two linked phenyl rings and various degrees of chlorination, were used widely in industry between the 1930s and 1970s,[1] and have since become ubiquitous and persistent environmental contaminants because of their resistance to chemical and thermal degradation.[2] The most common source of general population exposure is low-level contamination of the food chain; other known sources of exposure, like the workplace or contaminated sport fish, can produce higher body burdens.[3,4] PCBs are readily absorbed when ingested, can persist in fat tissue for years,[5,6] and can cross the placenta to affect the fetus.[7,8] Commercial grade PCBs per se are teratogenic, causing malformations in mice[9] and pigs,[10] and decreased intrauterine growth and survival in monkeys, rabbits, and rats.[11] In utero exposure has also increased motor activity and impaired learning in mice,[12–14] male rats,[15] and rhesus monkeys.[16–18]

Thus far, however, birth defects in human beings have been documented only after maternal exposure to a complex mixture of heat-degraded PCBs. Two such exposures have occurred, one in Japan (called Yusho, or "oil disease") in 1968[19] and one (Yu-cheng, the episode studied here) in Taiwan in 1979.[20] In both instances, PCBs were used as heat exchangers and thus were contaminated by their heat-degraded by-products, polychlorinated dibenzofurans, terphenyls, and quarterphenyls. These chemicals leaked into and became mixed with rice bran oil. Children born to women who had consumed the contaminated oil were found to have ectodermal defects, including hyperpigmentation, deformed nails, and natal teeth; intrauterine growth delay; and developmental delay.[21–23] The Japanese Yusho children were described as hypotonic, apathetic, and dull at 9 and 10 years old.[22]

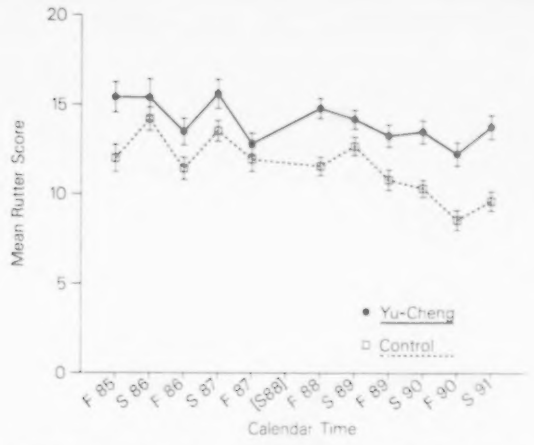
In the Taiwan incident, more than 2000 people who had consumed the contaminated oil for up to about 9 months between June 1978 and October 1979 were included in the Yu-cheng registry.[20] In the first 3 years after the outbreak, babies born to exposed women had a high infant mortality rate.[20] Surviving children were found to have ectodermal defects, developmental delay, more behavioral problems, and higher activity levels.[23–33] Followed from 1985 to 1991 with cognitive and behavioral tests, these children were also found to have poorer cognitive development up to 7 years of age.[28]

We examine here the behavior and activity level of the children born to exposed women. We explore both the relationship of the behavior and activity scores to the interval between outbreak and year of birth, and the possibility that the effect of exposure diminishes as the children age. We also describe the re-

Yung-Cheng Joseph Chen and Chen-Chin Hsu are with the Department of Psychiatry at the National Cheng Kung University Medical College, Tainan, Taiwan. Mei-Lin M. Yu, Beth C. Gladen, and Walter J. Rogan are with the Division of Intramural Research, National Institute of Environmental Health Sciences, Research Triangle Park, NC.

Requests for reprints should be sent to Walter J. Rogan, MD, MPH, MD A3-05, National Institute of Environmental Health Sciences, PO Box 12233, Research Triangle Park, NC 27709.

This paper was accepted August 23, 1993.



Note. F = fall, S = spring; behavior levels were not evaluated in spring 1988.

FIGURE 1—Mean Rutter scores by calendar time, Yu-cheng and control children.

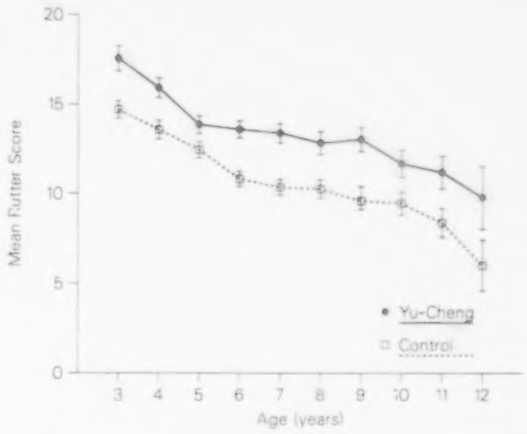


FIGURE 2—Mean Rutter scores by age, Yu-cheng and control children.

lationship of the behavior and activity scores to chemical, physical, and cognitive findings.

Methods

Study Subjects

The procedure used to identify the study children has been described in detail elsewhere.[23] Briefly, 132 living children, born to exposed women between June 1978 and February 1985, were identified. These children are referred to as "Yu-cheng children." One hundred and eighteen of them participated in a 6-year cognitive and behavioral follow-up study. For each Yu-cheng child, a control matched on age (within 15 days for those under 1 year, and within 1 month for older children), sex, neighborhood, maternal age (within 3 years), and socioeconomic status was selected for the follow-up study.

Behavioral and Activity Measurement Scales

There was no validated Chinese instrument for assessing children's behavior in 1985. Our group has clinical experience with behavioral disorders in Chinese children, and the items on Rutter's Child Behavior Scale A,[34] which had been validated in British children aged 9 to 12, appeared to be culture fair. The scale is a screening instrument to identify children likely to show some emotional or behavioral disorder; the three sections concern health problems, habits, and behaviors. The questionnaire was modified so that all items allowed a "not applicable" answer, and the parents used this option for younger children in whom one would expect that some of the behaviors asked about could not have occurred. There was a strong relationship between the percentage of not applicable answers and age: 21% for children under age 1, 11% for ages 1 to 2, 4.5% for ages 3 to 4, and less than 1% for ages 5 and over. Chinese parents reported a much higher frequency of behavioral problems in all children, including controls, than one would expect from the results in British children: 48% of the Yu-cheng children and 25.5% of the control children aged 9 to 12 years met Rutter's criterion for a behavioral disorder, compared with 15.1% and 8.1% of the British boys and girls, respectively, of the same ages. Nevertheless, it seemed justified to use the Rutter scale from age 3 years on because less than 5% of the questions were not applicable to children outside the 9- to 12-year age group, and because there were simultaneous control children and so the "normal" values for the instrument did not have to be used. However, it should be borne in mind that it is not known whether these higher scores in the Yu-cheng children, which represent more behavioral problems, are associated with the psychiatric diagnoses with which they are associated in British children. The test-retest reliability was assessed on 30 mothers with a 2-week interval between ratings, and the correlation between the scores on the two occasions was .82.[35]

A modified Werry-Weiss-Peters Activity Scale,[36] which had been validated in Chinese children aged 3 to 11,[37] was used to evaluate the children's activity level. Both behavior and activity questionnaires were filled out by a parent with instructions from an interviewer. The matched control was always evaluated on the same day as the Yu-cheng child. The evaluations were conducted twice a year between fall 1985 and spring 1991, except for spring 1988.

Data Analysis

Paired *t* tests were used to compare the mean Rutter behavior or activity scores between Yu-cheng and control children at each time point or age. Student's *t* tests were used to compare the mean scores of Yu-cheng children who had physical stigmata at birth or in the 1985 survey with the scores of those who did not, and to compare the scores of children who had detectable PCBs with the scores of those who did not. Simple regression analyses were used to test the relationship of the scores to maternal PCB values and IQ score.

Results

One hundred and fifteen of the original 118 Yu-cheng children com-

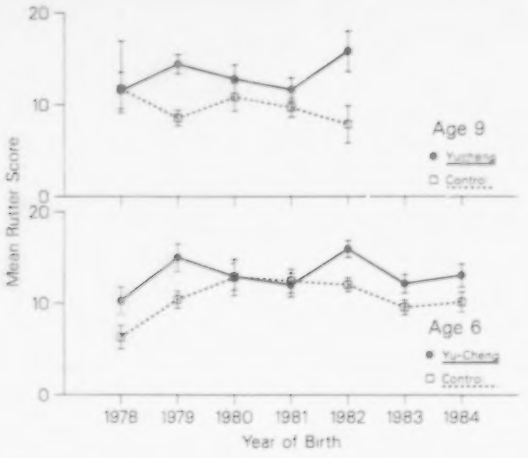


FIGURE 3—Mean Rutter scores at ages 6 and 9 by year of birth, Yu-cheng and control children.

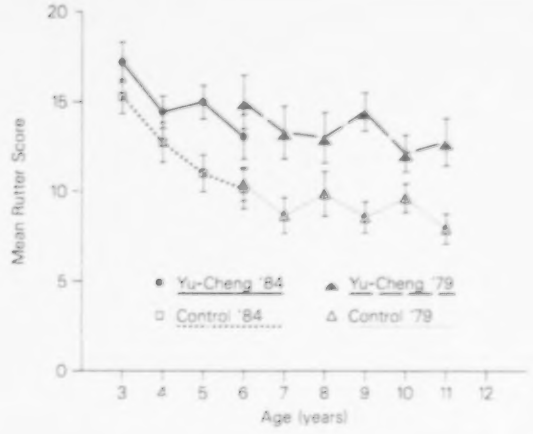


FIGURE 4—Mean Rutter scores for 1979 and 1984 birth cohorts, by age, Yu-cheng and control children.

pleted the 6-year follow-up study, as did their matched controls. The other three Yu-cheng children were removed from the study by their parents in early test rounds, and their corresponding controls were not evaluated further.

Because the children were born over a 7-year period, they were of different ages in any one round of follow-up. When children of different ages were combined and behavior scores analyzed by test time, the Yu-cheng children were found to have scored 7% to 43% (mean = 23%) higher than the control children on the Rutter scale at every time point (Figure 1); all differences were statistically significant except for those in spring 1986 and fall 1987. If analysis was done by age and the scores from different time points were combined, the Yu-cheng children scored 11% to 63% (mean = 28%) higher than their controls at each age (Figure 2); all differences were statistically significant except those for the 12-year-olds, for which there were only 17 pairs.

To examine whether children born longer after the exposure are less affected, the children were separated by year of birth. Figure 3 shows mean Rutter scores for all children aged 6 and 9 arrayed by year of birth. Except for the 1978 birth cohort at age 9 and the 1980 and 1981 birth cohorts at age 6, Yu-cheng children consistently scored higher than their controls. There was no consistent trend toward decreased differences in scores of Yu-cheng and control children as the interval between the exposure and year of birth increased. Pictures at other ages are similar.

To see whether the behavioral abnormality in the children persists as they grow, a given birth cohort was examined over time. The Rutter scores of the 1979 and the 1984 birth cohorts, arrayed by age, are shown in Figure 4. The difference between the Yu-cheng and the control children did not decrease as the children grew older.

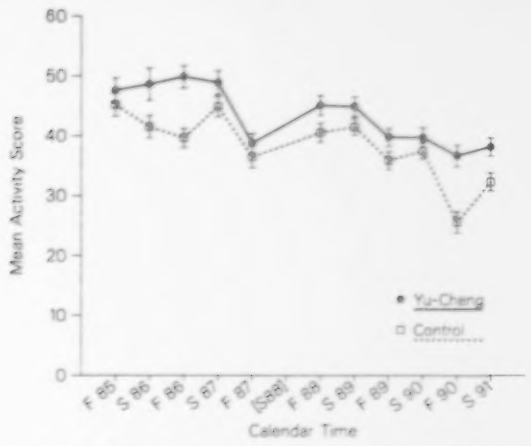
A similar but weaker picture was seen for the activity score. At each round, Yu-cheng children had a mean activity score 5% to 44% higher than the control children (Figure 5), and the differences were statistically significant at seven time points; Yu-cheng children scored 8% to 53% higher than their controls at each age (Figure 6), and the differences were significant in six age groups. The average differences across calendar time or across age were both highly significant.

There is information on physical findings, cognitive scores, and PCB levels of the Yu-cheng children from previous studies.[23,25,28,38] When the Rutter and activity scores of Yu-cheng children with various historical or physical stigmata were compared with the scores of those children without the stigmata, there did not appear to be any consistent trend. Children with physical signs had a higher mean score at some ages and a lower score at others (Table 1). Cognitive scores, PCB detectability, maternal serum PCB levels, and breast-feeding mode did not have consistent relationships to Rutter or activity scores, either.

Discussion

The data show that in utero exposure to PCBs and their heat-degraded contaminants is associated with mildly disordered behavior and increased activity level. The effects persisted as the children aged, and they appeared to be similar in children born up to 6 years after the mother was exposed. This is the first study to report the long-term effects of in utero exposure to heat-degraded PCBs on children's behavior and activity level.

Some of the children still have visible Yu-cheng–related physical stigmata, such as abnormal patterns of pigmentation, hypoplastic nails, and teeth that are very susceptible to caries. Since both the behavior questionnaire and the activity checklist were filled out by a parent, the perception of the parent that his or her child is "abnormal" may contribute to the differences in reported behavior and activity. So far, we have not seen consistent relationships between either cognitive[38] or behavioral function in the children and the presence of the physical stigmata of the syndrome. In general, the physical signs are subtle, and among children old enough to attend school, the results of teacher ratings on each child's maladaptive behavior and activity level are thus far



Note. F = fall, S = spring; activity levels were not evaluated in spring 1988.

FIGURE 5—Mean activity scores by calendar time, Yu-cheng and control children.

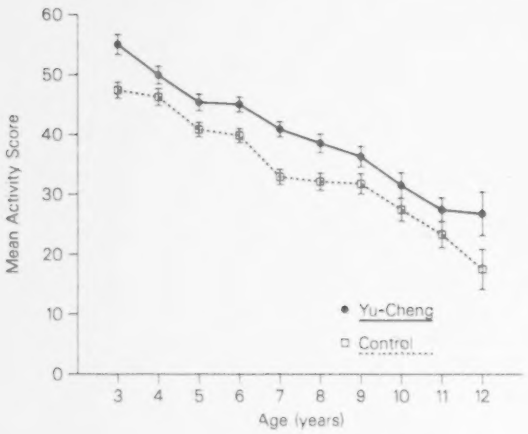


FIGURE 6—Mean activity scores by age, Yu-cheng and control children.

consistent with the parent's rating.[26,29,32,33] We believe the teachers are less likely to be aware of the child's exposure status and are thus not sensitized to the child's behavior. It is also possible that the perception that the child is damaged, regardless of whether there are physical signs, may in fact affect a parent's attitude in rearing the child and produce the abnormal behavior.

Yu-cheng children had persistent cognitive deficits up to 7 years of age.[28] The observed association between in utero PCB exposure and persistent behavior and activity level alteration may be secondary to the cognitive developmental delay. Our data did not show consistent and significant associations between cognitive scores and behavioral scores for the Yu-cheng children. This suggests that the developmental delay observed in these children did not produce secondary behavioral problems.

The finding of no consistent relationships between behavior/activity scores and physical findings, cognitive scores, and serum PCB levels is consistent with our previous report,[38] in which we found inconsistent relationships between cognitive developmental scores and physical stigmata and serum PCB levels. It is possible that our measures of dose did not accurately reflect the children's in utero exposure; that the congeners responsible for the physical stigmata may not have been those responsible for the cognitive delay or the disordered behavior; or that different mechanisms were involved in the physical stigmata, cognitive delay, and disordered behavior.

We matched the control children with the Yu-cheng children on age, sex, neighborhood, maternal age, and socioeconomic status. None of the exposed families had any psychiatric history, none of the exposed mothers had ever smoked, and only two exposed mothers had ever drunk regularly. However, other factors that we did not measure could have caused residual confounding; these factors may include parity, temperament of the mother, child-rearing attitudes or behaviors, and family and parental stresses. There is, however, no a priori reason to believe that control families differed in stress levels, attitudes, or illnesses other than those caused by exposure.

The Japanese Yusho children were described as hypotonic, apathetic, and dull at 9 and 10 years old.[22] However, since these children were not randomly selected and no controls were compared, the observation may not represent the whole picture of Yusho children.

Behavior and activity levels were evaluated in two US cohorts—one in Michigan, the other in North Carolina—exposed perinatally to background levels of PCBs. In North Carolina, using the information abstracted from report cards, researchers found no significant relationship between PCB exposure, either prenatally or through breast milk, and the child's work habit or conduct grades.[39] The study also found no association between exposure and hyperactivity reported by parents.[39] In the Michigan study, which used the reports from mothers and examiners when the children were 4 years of age, a composite activity rating was not related to prenatal PCB exposure.[40] The negative results in the two US cohorts were probably due to the low dose of PCBs and the virtual absence of polychlorinated dibenzofurans in the US exposures.

The Michigan study did find reduced activity among children with higher 4-year serum PCB levels.[40] These PCB levels were related to the children's breast-feeding history and not to their prenatal exposure, and the effect was seen only in children who had been breast-fed for more than 1 year and whose mothers had above average breast milk PCB levels. Breast-feeding has the potential to transfer much more of these chemicals to the child than does transplacental exposure. The Michigan findings were consistent with findings from two laboratory studies that found that postnatal exposure to PCBs reduced the activity level.[41–43] In our study, breast-feeding was not found to have a consistent association with the child's activity level among the Yu-cheng children. However, Yu-cheng mothers were advised not to breast-feed their children, and only 30 did.

Behavioral alterations appear in animals exposed perinatally to PCBs.[44]

The most consistent finding has been hyperactivity among animals—namely, mice, male rats, and rhesus monkeys—exposed in utero.[12-18] One group of rhesus monkeys exposed perinatally were hyperactive at 6 and 12 months of age but were hypoactive at 44 months.[16] The authors suggest that the altered behavior was a delayed and possibly permanent effect of the early exposure.

The continued mildly disordered behavior in these children is likely due to the persistence of the chemicals in their mothers, which resulted in the child's in utero exposure long after the mother's exposure to the chemicals ceased. Even though the levels in the mothers were declining, there appeared to be sufficient agent to be toxic for at least 6 years after the exposure. The persistent effect observed as the children aged may have been due to either continued internal exposure to PCBs/polychlorinated dibenzofurans stored in the body or permanent structural changes.

The biological mechanism that causes behavioral toxicity is unclear. It has been suggested that the PCB congeners responsible for neurochemical changes are different from the dioxin-like coplanar congeners that bind the Ah receptor and cause immunotoxicity, enzyme induction, and possible carcinogenesis.[45] After exposure of intact pigtail macaques (*Macaca nemestrina*) to two PCB mixtures, ortho- and ortho-, para-substituted congeners were found in the nigra-striatal system and the hippocampus, and the presence of these congeners correlated with decreases in the dopamine content in these brain regions.[45,46] Thus, these orthosubstituted nonplanar congeners may be responsible for the observed changes in in vivo neurochemistry. This suggestion is supported by the observation that there were no consistent relationships between behavioral scores and physical findings in the Yu-cheng children. One other possible mechanism is that the planar dioxin-like congeners alter brain steroid levels during development, which changes brain organization and the regulation of dopaminergic systems.[45]

The findings of mildly yet persistently disordered behavior and higher activity levels among the Yu-cheng children in this study is of public health significance. Even though PCBs have not been manufactured since the late 1970s, all PCB-containing equipment will have to be disposed of in some way. There will continue to be a risk of exposure to workers and the general population through accidents, improper disposal practices, transport of equipment containing huge amounts of PCBs, and contaminated animals and food. Prenatal exposure can cause activity alteration at a level that does not cause obvious clinical symptoms. The Yu-cheng children are the only well-documented cohort with sufficient in utero exposure to cause obvious toxicity. The behavioral and activity findings from this cohort may help us understand the toxicity that background exposure may cause. We will continue to follow these children and evaluate their health outcomes to understand the disease process. □

TABLE 1—Relationship between Physical, Cognitive, and Chemical Findings and Rutter Scores, Yu-cheng Children

	4 Years of Age		6 Years of Age		9 Years of Age	
	Mean ± SE	(n)	Mean ± SE	(n)	Mean ± SE	(n)
Neonatal eyelid swelling						
Yes	18.9 ± 1.7	(26)*	13.4 ± 1.3	(37)	13.6 ± 1.3	(19)
No	15.4 ± 0.5	(117)	13.8 ± 0.5	(141)	12.8 ± 0.8	(62)
Neonatal hyperpigmentation						
Yes	17.7 ± 1.0	(57)**	13.4 ± 0.7	(75)	13.3 ± 1.0	(44)
No	15.0 ± 0.6	(86)	13.9 ± 0.7	(103)	12.7 ± 0.9	(37)
Neonatal deformed/small nails						
Yes	19.1 ± 1.6	(32)**	12.1 ± 0.8	(44)**	14.2 ± 1.1	(24)
No	15.2 ± 0.5	(111)	14.2 ± 0.06	(134)	12.5 ± 0.8	(57)
History of bronchitis/pneumonia in first 6 months						
Yes	18.5 ± 1.5	(26)**	14.2 ± 1.0	(43)	13.8 ± 1.1	(24)
No	15.5 ± 0.6	(117)	13.5 ± 0.5	(135)	12.7 ± 0.8	(57)
History of toenail deformity						
Yes	18.0 ± 1.3	(39)*	13.6 ± 0.8	(51)	14.1 ± 1.1	(28)
No	15.4 ± 0.6	(102)	13.8 ± 0.6	(125)	12.4 ± 0.8	(53)
Deformed/pigmented toenails at initial examination						
Yes	15.9 ± 0.7	(90)	12.7 ± 0.6	(114)**	13.0 ± 0.8	(59)
No	16.4 ± 0.7	(53)	15.4 ± 0.8	(64)	13.2 ± 1.1	(22)
Cognitive score[a]						
< 100	16.7 ± 0.8	(51)**	13.9 ± 0.8	(77)	13.2 ± 1.2	(26)
≥ 100	13.8 ± 0.9	(35)	13.4 ± 1.5	(19)	12.5 ± 1.7	(14)
Detectable child's serum PCBs						
Yes	17.1 ± 1.2	(14)	11.2 ± 1.2	(23)**	13.3 ± 1.9	(13)
No	16.7 ± 1.3	(11)	18.3 ± 2.0	(11)	14.5 ± 4.7	(4)
Maternal serum PCB level						
≤ 25 ppb[b]	17.1 ± 0.9	(72)*	13.5 ± 0.6	(88)	14.6 ± 1.0	(31)**
> 25 ppb	14.9 ± 0.7	(61)	13.5 ± 0.8	(77)	11.1 ± 0.9	(42)
Feeding mode						
Breast-fed	18.7 ± 1.8	(19)*	13.3 ± 0.8	(46)	13.6 ± 1.0	(29)
Bottle-fed	15.5 ± 0.6	(124)	13.9 ± 0.6	(132)	13.2 ± 0.8	(51)

[a]Stanford-Binet test for age 4, and Webster Intelligence Scale for Children, Revised (WISC-R) for ages 6 and 9.
[b]Median value.
*Groups differ at $P < .10$ by t test; **groups differ at $P < .05$ by t test.

Acknowledgments

This research was supported by grants NSC74-0301-H002-29, NSC76-0301-H002-09, NSC77-0301-H002-24, NSC78-0301-H006-03, NSC79-0301-H006-02, and NSC80-0301-H006-07 from the National Science Council, Taiwan, Republic of China.

We are indebted to the staff at the Taiwan Provincial Health Department and

local health stations for their support and cooperation; to N. Beth Ragan for assistance in graphics; and to Drs Ruth Little and Eric Johnson for their constructive suggestions about this manuscript.

References

1. de Voogt P, Brinkman UAT. Production, properties and usage of polychlorinated biphenyls. In: Kimbrough RD, Jensen AA, eds. *Topics in Environmental Health:* vol 4. *Halogenated Biphenyls, Terphenyls, Naphthalenes, Dibenzodioxins, and Related Products.* 2nd fully rev ed. Amsterdam, the Netherlands: Elsevier; 1989:3–45.
2. Anderson HA. General population exposure to environmental concentrations of halogenated biphenyls. In: Kimbrough RD, Jensen AA, eds. *Topics in Environmental Health:* vol 4. *Halogenated Biphenyls, Terphenyls, Naphthalenes, Dibenzodioxins, and Related Products.* 2nd fully rev ed. Amsterdam, the Netherlands: Elsevier; 1989:325–344.
3. Maroni M, Colombi A, Cantoni S, Ferioli E, Foa V. Occupational exposure to polychlorinated biphenyls in electrical workers: I. environmental and blood polychlorinated biphenyls concentrations. *Br J Ind Med.* 1981;38:49–54.
4. Humphrey HEB. Population studies of PCBs in Michigan residents. In: D'Itri FM, Kamrin MA, eds. *PCBs: Human and Environmental Hazards.* Boston, Mass: Butterworth; 1983:293–310.
5. Safe S. Polyhalogenated aromatics: uptake, disposition and metabolism. In: Kimbrough RD, Jensen AA, eds. *Topics in Environmental Health:* vol 4. *Halogenated Biphenyls, Terphenyls, Naphthalenes, Dibenzodioxins, and Related Products.* 2nd fully rev ed. Amsterdam, the Netherlands: Elsevier; 1989:131–159.
6. Steele G, Stehr-Green P, Welty E. Estimates of the biologic half-life of polychlorinated biphenyls in human serum. *N Engl J Med.* 1986;314:926–927. Letter.
7. Kodama H, Ota H. Transfer of polychlorinated biphenyls to infants from their mothers. *Arch Environ Health.* 1980;35:95–100.
8. Masuda Y, Kagawa R, Kuroki H, et al. Transfer of polychlorinated biphenyls from mothers to foetuses and infants. *Food Cosmetic Toxicol.* 1978;16:543–546.
9. Marks TA, Kimmel GL, Staples RE. Influence of symmetrical polychlorinated biphenyl isomers on embryo and fetal development in mice: I. teratogenicity of 3,3',4,4',5,5'-hexachlorobiphenyl. *Toxicol Appl Pharmacol.* 1981;61:269–276.
10. Earl FL, Couvillion JL, Van Loon EJ. The reproductive effects of PCB 1254 in beagle dogs and miniature swine. *Toxicol Appl Pharmacol.* 1974;29:104. Abstract.
11. Morrissey RE, Schwetz BA. Reproductive and developmental toxicity in animals. In: Kimbrough RD, Jensen AA, eds. *Topics in Environmental Health:* vol 4. *Halogenated Biphenyls, Terphenyls, Naphthalenes, Dibenzodioxins, and Related Products.* 2nd fully rev ed. Amsterdam, the Netherlands: Elsevier; 1989:195–225.
12. Chou SM, Miike T, Payne WM, Davis GJ. Neuropathology of "spinning syndrome" induced by prenatal intoxication with a PCB in mice. *Ann N Y Acad Sci.* 1979;320:373–395.
13. Storm JE, Hart JL, Smith RF. Behavior of mice after pre- and postnatal exposure to Arochlor 1254. *Neurobehav Toxicol Teratol.* 1981;3:5–9.
14. Tilson HA, Davis GJ, McLachlan JA, Lucier GW. The effects of polychlorinated biphenyls given prenatally on the neurobehavioral development of mice. *Environ Res.* 1979;18:466–474.
15. Shiota K. Postnatal behavioral effects of prenatal treatment with PCBs (polychlorinated biphenyls) in rats. *Okajimas Folia Anat Jpn.* 1976;53:105–114.
16. Bowman RE, Heironimus MP. Hypoactivity in adolescent monkeys perinatally exposed to PCBs and hyperactive as juveniles. *Neurobehav Toxicol Teratol.* 1981;3:15–18.
17. Bowman RE, Heironimus MP, Allen JR. Correlation of PCB body burden with behavioral toxicology in monkeys. *Pharmacol Biochem Behav.* 1978;9:49–56.
18. Bowman RE, Heironimus MP, Barsotti DA. Locomotor hyperactivity in PCB-exposed rhesus monkeys. *Neurotoxicology.* 1981;2:251–268.
19. Kuratsune M, Yoshimura T, Matsuzaka J, Yamaguchi A. Epidemiologic study on Yusho, a poisoning caused by ingestion of rice oil contaminated with a commercial brand of polychlorinated biphenyls. *Environ Health Perspect.* 1972;1:119–128.
20. Hsu S-T, Ma C-I, Hsu SK-H, et al. Discovery and epidemiology of PCB poisoning in Taiwan: a four-year followup. *Environ Health Perspect.* 1985;59:5–10.
21. Funatsu I, Yamashita F, Ito Y, et al. Polychlorbiphenyls (PCB) induced fetopathy: I. clinical observation. *Kurume Med J.* 1972;19:43–51.
22. Harada M. Intrauterine poisoning: clinical and epidemiological studies of the problem. *Bull Inst Constitutional Med Kumamoto Univ.* 1976;25(suppl):1–69.
23. Rogan WJ, Gladen BC, Hung K-L, et al. Congenital poisoning by polychlorinated biphenyls and their contaminants in Taiwan. *Science.* 1988;241:334–336.
24. Gladen BC, Rogan WJ, Ragan NB, Spierto FW. Urinary porphyrins in children exposed transplacentally to polyhalogenated aromatics in Taiwan. *Arch Environ Health.* 1988;43:54–58 (errata 348).
25. Gladen BC, Taylor JS, Wu Y-C, Ragan NB, Rogan WJ, Hsu C-C. Dermatologic findings in children exposed transplacentally to heat-degraded polychlorinated biphenyls in Taiwan. *Br J Dermatol.* 1990;122:799–808.
26. Chen C-C, Hsu C-C, Yeh T-L, Lin S-C, Duann Y-H. A six-year follow-up study of intellectual and behavioral development of Yu-cheng children: cross-sectional findings of the second field work study [in Chinese; English summary]. *Chin Psychiatry.* 1988;2:257–266.
27. Chen Y-C, Guo Y-L, Hsu C-C. The cognitive and behavioral development of children prenatally exposed to polychlorinated biphenyls and contaminants: sixth-year fieldwork report. *Chin Psychiatry.* 1992;6:116–125.
28. Chen Y-C, Guo Y-L, Hsu C-C, Rogan WJ. Cognitive development of Yu-cheng ("oil disease") children prenatally exposed to heat-degraded PCBs. *JAMA.* 1992;268:3213–3218.
29. Chen Y-C, Hsu C-C, Soong W-T, et al. A six-year follow-up study of intellectual and behavioral development of Yu-cheng children: findings during the 3rd year of field work [in Chinese; English summary]. *Chin Psychiatry.* 1989;3:89–98.
30. Chen Y-C, Yeh T-L, Hsu C-C. A six-year follow-up study on the intellectual and behavioral development of Yu-cheng (oil disease) children: findings for the fifth year of field work [in Chinese; English summary]. *Chin Psychiatry.* 1990;4:40–51.
31. Hsu C-C, Chen T-C, Soong W-T, et al. A six-year follow-up study of intellectual and behavioral development of Yu-cheng children: cross-sectional findings of the first field work [in Chinese; English summary]. *Chin Psychiatry.* 1988;2:27–40.
32. Hsu C-C, Chen Y-C, Soong W-T, Ko H-C. A six-year follow-up study of intellectual and behavioral development of Yu-cheng (oil disease) children: cross-sectional findings of the fourth year of field work [in Chinese]. *Chin Psychiatry.* 1989;3(suppl 1):101–111.
33. Yeh T-L, Hsu C-C, Chen C-C, et al. A six-year follow-up study of intellectual and behavioral development of Yu-cheng children: findings during the second year of field work [in Chinese; English summary]. *Chin Psychiatry.* 1988;2:172–185.
34. Rutter M. Appendix 6: a children's behavior questionnaire for completion by parents. In: Rutter M, Tizard J, Whitmore K, eds. *Education, Health, and Behavior: Psychological and Medical Study of Childhood Development.* New York, NY: Wiley; 1970:412–421.
35. Hsu C-C. A six-year follow-up study of intellectual and behavioral development of Yu-cheng children: the first year annual report to the National Science Council [in Chinese; English summary]. Taipei, Taiwan: National Taiwan University Hospital; 1986.
36. Werry JS. Developmental hyperactivity. *Pediatr Clin North Am.* 1968;15:581–599.
37. Hsu C-C, Lin C-C, Kuo H-L. A preliminary study of the activity level of Chinese children aging three to eleven years [in Chinese; English summary]. *Natl Sci Counc Monthly, ROC.* 1982;10:363–381.
38. Yu M-L, Hsu C-C, Gladen BC, Rogan WJ. In utero PCB/PCDF exposure: relation of developmental delay to dysmorphology and dose. *Neurotoxicol Teratol.* 1991;13:195–202.
39. Rogan WJ, Gladen BC. Neurotoxicology of PCBs and related compounds. *Neurotoxicology.* 1992;13:27–35.
40. Jacobson JL, Jacobson SW, Humphrey HEB. Effects of exposure to PCBs and related compounds on growth and activity of children. *Neurotoxicol Teratol.* 1990;12:319–326.
41. Koja T, Fujisaki T, Shimizu T, Kishita C, Fukuda T. Changes of gross behavior with polychlorinated biphenyls (PCB) in immature rats [in Japanese]. *Kagoshima Daigaka Igaka Zasshi.* 1978;30:377–381. Cited in Tilson et al.[44]
42. Koja T, Kishita C, Shimizu T, Fujisaki T,

Kitazoro M, Fukuda T. Effects of polychlorinated biphenyls (PCB) on the gross behavior of immature rats and the influence of drugs upon them [in Japanese]. *Kagoshima Daigaka Igaka Zasshi.* 1979;31: 315–319. Cited in Tilson et al.[44]

43. Pantaleoni G, Fanini D, Sponta AM, Palumbo G, Giorgi R, Adams PM. Effects of maternal exposure to polychlorobiphenyls (PCBs) on F1 generation behavior in the rat. *Fundam Appl Toxicol.* 1988;11:440–449.
44. Tilson HA, Jacobson JL, Rogan WJ. Polychlorinated biphenyls and the developing nervous system: cross-species comparisons. *Neurotoxicol Teratol.* 1990;12: 239–248.
45. Seegal RF, Bush B, Shain W. Lightly chlorinated ortho-substituted PCB congeners decrease dopamine in nonhuman primate brain and in tissue culture. *Toxicol Appl Pharmacol.* 1990;106:136–144.
46. Seegal RF, Bush B, Brosch KO. Comparison of the effects of Aroclors 1016 and 1260 on non-human primate catecholamine function. *Toxicology.* 1991;66:145–163.

Air Recirculation and Sick Building Syndrome: A Blinded Crossover Trial

Jouni J. K. Jaakkola, MD, DSc, Pekka Tuomaala, TechLic, and Olli Seppänen, TechLic

ABSTRACT

Objective. This study tested the hypothesis that recirculated air in mechanically ventilated buildings causes symptoms commonly referred to as the sick building syndrome and perceptions of poor indoor air quality.

Methods. A blinded, four-period crossover trial was carried out in two identical buildings, contrasting 70% return air (index phase) with 0% of return air (reference phase). Each period lasted 1 workweek. The study population comprised 75 workers who had reported symptoms related to the work environment or perceptions of poor indoor air quality. Participants reported their ratings of symptoms, their perceptions, and related information in a daily diary. The outcome criteria included aggregative symptom scores for mucosal irritation, skin reaction, allergic reaction, and general symptoms formed of ratings of component symptoms. Perceptions of unpleasant odor, stuffiness, or dustiness were additional outcome criteria.

Results. All 75 participants returned their diaries. For no symptoms did the scores differ between the two phases more than could be expected by chance. Mean rating of unpleasant odor was significantly smaller during the index phase, but mean ratings of dustiness and stuffiness did not differ materially between the two phases.

Conclusions. Our results suggest that 70% recirculated air, when accompanied by an adequate intake of outdoor air, can be used without causing adverse effects. (*Am J Public Health.* 1994;84:422–428)

Introduction

Office workers in North America and Europe have been reported to have symptoms and other health problems attributed to the work environment.[1–12] These symptoms are nonspecific and occur in any population, but several studies indicate that they are more common among the occupants of certain buildings, often referred to as "sick buildings." Thus, the recurrence of these symptoms has given rise to the concept of the sick building syndrome. According to a World Health Organization working group, this syndrome is characterized by eye, nose, and throat irritation; a sensation of dry mucous membranes and skin; erythema; mental fatigue; headache; a high frequency of airway infections and cough; hoarseness; wheezing, itching, and nonspecific hypersensitivity; nausea, and dizziness.[13] There have been several other attempts to define the syndrome as well.[3,14–17]

The indoor environment in a confined space is a complex, dynamic combination of physical, chemical, and biological factors that may affect human health and prompt physical reactions. There is evidence that volatile organic compounds commonly found in office air can cause symptoms typical of the sick building syndrome.[18,19] Exposure to environmental tobacco smoke, which consists of various chemical substances and particulates, has also been related to these symptoms.[4,6,20] Studies in which people subjected to spaces with wall-to-wall carpets,[21] fleecy material,[10] or textile surface material[22] reported more symptoms than people in similar spaces without those materials have provided indirect evidence of the importance of biological particles. In addition, indoor air temperature[5,23] above 22°C and relative humidity[24] below 25% have also been associated with an excess of sick building syndrome symptoms in cross-sectional studies. A six-period crossover trial[25] among office workers was able to show that symptoms common in low relative humidity (20% to 25%) can be prevented by modest air humidification (raising relative humidity to 30% to 35%).

In addition to the quality of outdoor air and emissions from the indoor environment and occupants, indoor air quality is affected by type and performance of heating, ventilating, and air-conditioning technology. In mechanical ventilation and air-conditioning systems, recirculation of air is used to control temperature and air distribution and to conserve energy. In this type of ventilation system, a part of the exhaust air from the rooms (return air) is recirculated back to the supply airflow, which is a mixture of return air and outdoor air (Figure 1). Air recirculation can be quantified in terms of the proportion of return air in the total airflow. Proportions of recirculated air as high as 80% to 90% are common in North America whereas, in Finland, they are usually between 30% and 70%.

Jouni J. K. Jaakkola is with the Department of Public Health at the University of Helsinki, Finland. Pekka Tuomaala is a research fellow and Olli Seppänen is with the Laboratory of Heating, Ventilating and Air Conditioning, Helsinki University of Technology.

Requests for reprints should be sent to Jouni J. K. Jaakkola, MD, DSc, Department of Public Health, PO Box 21 (Haartmaninkatu 3), SF-00014 University of Helsinki, Finland.

This paper was accepted August 10, 1993.

Editor's Note. See related editorial by Mendell and Fine (p 346) in this issue.

The rate of the total and outdoor air supply are other important parameters when studying the effects of air recirculation. It is evident when outdoor airflows are too low, indoor pollutants can accumulate in concentrations that can cause sick building syndrome symptoms. In 1987, a Nordic research meeting considered the use of air recirculation to be a "risk solution" and recommended against it in public buildings.[26] This recommendation was justified by the risk of spreading indoor air pollutants and by the prevailing poor technology and maintenance of systems involving recirculation.

We carried out a blinded, four-period crossover trial to test the hypothesis that the use of recirculated air in mechanically ventilated office buildings with sufficient intake of outdoor air (exceeding the regulatory recommendations) and without any unusual internal sources of indoor air pollution causes mucosal irritation, skin reactions, allergic reactions, and general symptoms (commonly known collectively as the sick building syndrome), as well as perceptions of unpleasant odor, stuffiness, or dustiness.

Methods

Buildings and Study Population

Two identical office buildings in Kilo, 15 km from the center of Helsinki, were selected as suitable for our study. Constructed in 1974, these buildings have eight stories and are 72 m long, 18 m wide, and 35 000 m^3 in volume. The structure of each is concrete, and the windows can be opened. Each building has a central mechanical ventilation system, hot water radiators for heating, and two identical air handling units, which are operated only during office hours, from 7 AM to 5 PM. The proportion of return air, ordinarily 30% to 40% during the heating season, is selected by adjusting the blades of the dampers.

From the source population—the 470 workers in the two buildings—all those who met the following eligibility criteria were targeted for recruitment to the study population: (1) experience, during the previous 12 months, of symptoms and/or perceptions of poor indoor air quality, which were attributed to the work environment generally or to the indoor air quality specifically; (2) work in an assigned room on the average of at least 3 days a week; (3) no anticipated absence from work during the study period owing to vacation, trip, or other reason; (4) no doctor-diagnosed asthma; and (5) no regular exposure to environmental tobacco smoke in the office. The resultant study population—the 75 persons who fulfilled these criteria—is described in Table 1.

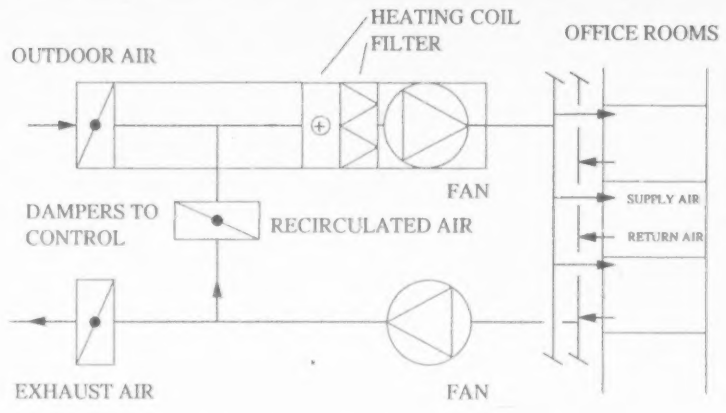


FIGURE 1—A schematic presentation of the air handling system used in the two buildings.

TABLE 1--Characteristics of the Study Population

	Building A (n = 37)		Building B (n = 38)		Total (n = 75)	
	No.	%	No.	%	No.	%
Age						
≤24	3	8	2	5	5	6
25–34	8	22	12	32	20	27
35–44	17	46	13	34	30	40
45–54	7	19	10	26	17	23
≥55	2	5	1	3	3	4
Gender						
Male	15	41	18	47	33	44
Female	22	59	20	53	42	56
Atopic eczema	5	14	8	21	13	17
Hay fever	12	33	17	45	29	39
Allergic conjunctivitis	8	22	11	29	19	25
Smoking						
Current	9	24	12	32	21	28
Ex	12	33	13	34	25	33
Never	16	43	13	34	29	39

Recirculation

We tested our hypothesis in a blinded four-period crossover trial in November-December 1988; each period consisted of 1 workweek. For the week just before the experiment, both buildings were ventilated without recirculation, with a total airflow of 20 liters per second (L/s) per person. During the first experimental week, based on random selection, one building (building A) was operated with 70% recirculation while no recirculation was used in the other building (building B). The operation of the recirculation system was then switched during the weekend so that this procedure was reversed during the second week. A similar crossover procedure was carried out two more times. Thus, each participant experienced two periods of exposure to recirculated air (index periods, index phase) and two

periods of no exposure (reference periods, reference phase).

The total supply airflow was kept constant at 20 L/s per person through the experiment. The system was designed to provide outdoor air at the rate of at least 4 L/s per person (the recommended minimum rate in Nordic countries during the study). No air humidification was used in the buildings. The participants were told that ventilation would be adjusted during the study, but neither the objective nor the phase of the study was revealed. Thus, the trial was blinded.

Outcome Criteria

Apart from particular symptoms per se, four outcome scores were defined, based on theories regarding potential health effects of air recirculation, and the reporting of corresponding symptoms (i.e., the symptoms that appear in each outcome score based on the theory) was used in forming them. With a partial overlap of elements, these scores were as follows:

1. *Mucosal irritation score*, based on scores for dryness, itching, or irritation of eyes; nasal dryness; nasal congestion ("stuffy nose"); and pharyngeal irritation
2. *Skin reaction score*, based on scores for dryness, itching, or irritation of skin, and rash
3. *Allergic reaction score*, based on scores for dryness, itching, or irritation of eyes; nasal congestion; nasal excretion ("runny nose"); sneezing; and cough
4. *General symptom score*, based on scores for headache and lethargy

Scores for the perception of unpleasant odor, stuffiness, or dustiness were additional outcome criteria.

Data Collection

During the trial, participants were asked to fill out a diary each day after work recording whether and to what extent they had experienced any of the outcome symptoms (structured answers: no = 0, mild = 1, moderate = 2, severe = 3) and had perceived unpleasant odor, stuffiness, or dustiness (scale from none = 0 to pervasive = 5). A sensation of dryness and of temperature was also requested (scale: all too humid/cold = 1, too humid/cold = 2, acceptable = 3, too dry/warm = 4, all too dry/warm = 5). Extraneous determinants of the outcomes were also inquired about daily. These included symptoms of the common cold (no symptoms = 0, cold and/or sore throat = 1, cold and/or sore throat and fever = 2); sensations of tobacco smoke in a nonsmoking room (scale from none = 0 to extensive odor = 3); the number of cigarettes smoked in the office by participant, roommates, and visitors; the time the window was open, and the reason for opening the window.

Air Measurements

Airflow, temperature, and relative humidity were measured on Tuesdays and Thursdays in all rooms during each period. Airflow was measured in the exhaust air register of each office using a calibrated Wallac anemometer with a relative error of less than 15%. Air temperature and relative humidity were measured with VAISALA HMI-31 capacitive sensors whose accuracy is ±0.3°C for temperature and ±2% for relative humidity. Supply airflow and proportion of return air were measured in the beginning and end of each period using the tracer gas method.

Data Reduction

The detailed data consisted of daily ratings (maximum 20 days) of the component symptoms, and perceptions of and structured answers to the questions on the extraneous factors. The outcome scores were calculated by adding the component symptom ratings and dividing the sum by the number of symptoms. For each participant, the means of the daily outcome scores and the ratings for the component symptoms and the perception of indoor air quality were calculated for both index and reference phases and the difference between these. Days in which less than 2 hours were spent in the office building were excluded. The subjects were also classified, for each item, according to those who reported more, the same number, or fewer symptoms/perceptions of poor indoor air quality during the index phase.

Statistical Methods

The primary hypothesis—that the use of recirculated air causes symptoms and perception of poor indoor air quality—was tested by comparing the means of the outcome criteria in the index, or recirculation, phase with those of the reference phase. In the primary analysis, differences between these means were assessed for statistical significance by using the paired t test, and the 95% confidence interval was calculated for the estimate.[27] The role of period effect was assessed in a general linear model using the SAS® computer package (procedure *proc glm*), as described by Senn.[27] The probability ratio between experiencing more and fewer symptoms/perceptions during the index phase as compared with the reference phase was also estimated. This was done by dividing the number of participants with more symptoms/perceptions during the index phase by the number of participants with fewer symptoms. The statistical significance was evaluated by means of McNemar's test, and the 95% confidence interval for the probability ratio was calculated by Miettinen's test-based method.[28] The primary analyses were carried out blinded to the two phases. The analyses were carried out first for the total study population and then for the subjects who had in the baseline questionnaire indicated having the corresponding symptom/perception during the previous 12 months.

Results

Experimental Conditions

The proportion of return air during the study period was achieved reasonably well according to the study design. It ranged from 69% to 71% during the index periods and from 1% to 6% during the reference periods (the result of leaky dampers). The total supply airflow in the 67 study rooms was somewhat lower during the index phase (mean = 20 L/s per person, range = 6.3–95.6, SD = 13.2) than during the reference phase (mean = 23 L/s per person, range = 6.0–77.2, SD = 12.8). Temperature was somewhat higher during the index phase (23.4°C, range = 21.9–24.8, SD = 0.6) than during the reference phase (23.1°C, range = 21.8–24.4, SD = 0.6). The relative humidity ranged from 23% to 28% during the first 3 weeks and from 14% to 20% during the fourth week. The mean relative humidity was 1% greater during the phase with air recirculation than during the phase without.

Participation

All 75 members of the study population returned their symptom diaries. Focus on the subjects who were in the office for at least 2 hours on at least 1 day in each of the two phases led to the deletion of three subjects.

Outcomes

The means of the symptom scores and the component symptom ratings during the index and reference phases are given in Table 2, together with the differences between them and their 95% confidence intervals based on intraindividual differences. The mean differences in the symptom scores did not differ from zero more than could be expected by chance. Of the single symptoms, the mean rating for sneezing was significantly lower during the index phase (0.24) than during the reference phase (0.33). The means of the outcome scores/ratings were higher in those who had corresponding symptoms during the previous 12 months, but there were no significant differences between the two phases.

Ratings of air for perceived poor quality were not higher during the index phase than during the reference phase (Table 2). On the contrary, the mean of unpleasant odor ratings was significantly lower during the index phase (0.19 vs 0.35).

The numbers of participants with more, fewer, and the same number of different symptoms and perceptions during the index phase as compared with the reference phase and the empirical probability ratio for having more symptoms/perceptions are given in Table 3. The number of subjects with more problems during the use of recirculated air did not differ significantly from the number of those with fewer problems, either in the total study population (Table 3) or in those with corresponding symptoms/perceptions during the previous 12 months.

Extraneous Factors

There were slightly more subjects with symptoms of the common cold during the reference period than during the index period (20 vs 23). The participants spent, on average, 8 minutes per day less in the office during the index phase than during the reference phase. Similarly, there were slight differences in smoking in the office (personal smoking per day: index = 0.52 cigarettes, reference = 0.60 cigarettes; other smoking per day: index = 0.22 cigarettes, reference = 0.24 cigarettes), perception of tobacco odor (index = 0.29, reference = 0.25), and opening of the windows (index = 11 minutes per day, reference = 13 minutes per day) during the two periods.

TABLE 2—Means of Symptom Scores and Ratings of Component Symptoms during Index and Reference Phases and Their Differences

Outcome	Index	Reference	Intraindividual Differences	
			Point Estimate	95% CI
Mucosal irritation score	0.42	0.44	−0.015	−0.066, +0.036
Eye symptoms[a]	0.22	0.24	−0.014	−0.067, +0.039
Nasal dryness	0.70	0.68	0.020	−0.081, +0.126
Nasal congestion[b]	0.51	0.50	0.004	−0.102, +0.110
Pharyngeal irritation	0.27	0.34	−0.070	−0.174, +0.034
Skin reaction score	0.15	0.16	−0.002	−0.024, +0.020
Skin symptoms[a]	0.50	0.50	−0.000	−0.065, +0.065
Rash	0.12	0.13	−0.008	−0.047, +0.031
Allergic reaction score	0.30	0.34	−0.034	−0.079, +0.011
Nasal excretion	0.25	0.29	−0.036	−0.111, +0.039
Sneezing	0.24	0.33	−0.090	−0.161, −0.019
Cough	0.18	0.24	−0.051	−0.129, +0.027
Eye symptoms[a]				
Nasal congestion[b]				
General symptom score	0.20	0.18	0.020	−0.023, +0.063
Headache	0.23	0.20	0.032	−0.023, +0.087
Lethargy	0.16	0.15	0.008	−0.055, +0.071
Perception of indoor air quality				
Unpleasant odor	0.19	0.35	−0.158	−0.289, −0.027
Stuffiness	0.84	0.84	−0.008	−0.188, +0.172
Dustiness	0.74	0.78	−0.037	−0.131, +0.057

[a]Dryness, itching, or irritation.
[b]Included in the allergic score.

TABLE 3—Probability Ratio between Having More and Fewer Symptoms or Discomfort during Index Phase than during Reference Phase

Outcome	Occurrence during 70% Air Recirculation				
	No. Subjects with More Symptoms	No. Subjects With Fewer Symptoms	No. Subjects with Equal Symptoms	Probability Ratio	
				Point Estimate	95% CI
Mucosal irritation score	25	30	17	0.83	0.49, 1.42
Eye symptoms[a]	13	19	40	0.68	0.33, 1.39
Nasal dryness	22	15	35	1.47	0.76, 2.83
Nasal congestion[b]	20	18	34	1.11	0.59, 2.10
Pharyngeal irritation	18	20	34	0.90	0.47, 1.72
Skin reaction score	21	21	30	1.00	0.58, 1.91
Skin symptoms[a]	20	18	34	1.11	0.59, 2.10
Rash	10	9	53	1.11	0.46, 2.70
Allergic reaction score	23	33	16	0.70	0.41, 1.18
Nasal excretion	14	19	40	0.74	0.38, 1.46
Sneezing	12	26	34	0.46	0.24, 0.90
Cough	7	20	45	0.35	0.15, 0.79
Eye symptoms[a]					
Nasal congestion[b]					
General symptom score	26	22	24	1.18	0.67, 2.08
Headache	25	16	31	1.56	0.84, 2.89
Lethargy	11	12	49	0.92	0.42, 2.00
Perception of indoor air quality					
Unpleasant odor	9	19	44	0.47	0.21, 1.05
Stuffiness	19	19	34	1.00	0.53, 1.89
Dustiness	13	16	43	0.81	0.34, 1.91

[a]Dryness, itching, or irritation.
[b]Included in the allergic score.

Discussion

The possibility that the use of recirculated air causes adverse health effects or discomfort became a public concern at a time when there was no empirical evidence to support or refute it. However, there were two theoretical models according to which the use of recirculated air could cause adverse effects. Berglund and Lindvall[29] suggested that the human sensory system uses a pattern recognition mechanism in the sensation of indoor air with complex environmental adaptation to the inhaled air. Use of a high proportion of recirculated air can lead to extreme homogenization of air, causing sensory confusion and strain on the system when it is trying to interpret the signals. According to the hypothesis, "sensory symptoms tied to 'sick buildings' of the irritant type"[25(p157)] could be related to the homogenization of air. In another theoretical model, which was based on the physical property of recirculation, indoor air pollutants from different point sources are circulated, causing low-level exposure to a mixture of chemical and biological pollutants, which can in turn cause mucosal irritation, skin and allergic reactions, and general symptoms. Thus, the use of air recirculation was hypothesized to produce conditions with adverse health effects; the methods to measure these conditions, however, are poor even with the best of technology.

The sick building syndrome has not been defined properly as a scientific concept, and so far no unifying mechanism has been postulated. We attempted to define the outcome criteria conceptually as biological reactions, mucosal irritation, and allergic reactions. We described the reactions operationally by the chosen symptom scores. The skin reaction score and general symptom score were also used as outcome criteria without any hypothesized mechanism; this was justified because these symptoms have also been attributed to the sick building syndrome.[1-4,8,13-15] General symptoms have also been addressed in earlier studies[8-10,29] without a hypothesized mechanism. The choice of symptoms related to a given mechanism is difficult, and there may be disagreement about the correctness of the choices. In our results, we also showed the occurrence of each component symptom to enable critical readers to evaluate the choice of symptoms or to choose their own combination.

The study population of 75 office workers was recruited from a source population of 470 office workers because they indicated having symptoms or perceived poor air quality related to the work environment at baseline. These subjects thus represented the most sensitive workers, those in whom the effect was most likely to be manifest. Further analyses were carried out focusing on those who had indicated having the corresponding symptom or perception related to the work environment during the previous 12 months when 30% to 40% of recirculated air had been used.

The experimental study design ensured the study's validity because intraindividual comparison of symptoms between different environmental conditions eliminated the potential confounding by personal characteristics. The exact purpose and the phase of the study were not revealed to the study subjects, and thus the information from the index and reference periods was comparable. The length of the periods, 1 week, was deemed adequate for postulated short-term effects. The weekends were considered, a priori, as sufficient washout periods to prevent the effect of exposure in one period to be carried over into a subsequent period. The period effect was found to be negligible in assessing the effect of air recirculation.

The amount of outdoor airflow is directly related to the capability of removing indoor air pollution from a given space. Keeping the total supply airflow constant and varying the proportion of recirculated air, as we did in our study, corresponds to the real-life situation. Our design purpose was to keep room temperature, relative humidity, and total supply airflow similar during the index and reference periods. However, the use of air recirculation had some minor effects on other indoor air factors: the room temperature increased, on average, by 0.4°C; the relative humidity increased by 1%; and the supply air intake rate decreased by 3 L/s per person. Room temperature above 22°C has been shown to increase the symptoms associated with the sick building syndrome.[5] According to an earlier study in Finland,[24,25] an increase in relative humidity in this range is likely to decrease the occurrence of eye and skin symptoms. In the marginal range, a decrease of supply airflow could increase indoor air pollution and thus the symptoms, but during the study the mean supply airflow was 21 L/s per person (minimum was 6 L/s per person) and thus the average decrease was less than 15%. During the index period, the mean outdoor airflow was 6 L/s per person (minimum was 2 L/s per person). In all, the differences in temperature, relative humidity, and supply airflow between the index and reference periods were small and thus were not likely to affect the results.

Uneven distribution of extraneous factors could have affected the validity of the study. To take this potential confounding into account, most important extraneous factors were recorded during the study period. Symptoms of the common cold were slightly more common during the reference phase, which may explain the observed excess in sneezing, nasal excretion, and cough. In all, the differences in the extraneous factors were very small and thus were unlikely to detract from the validity of the study.

The use of 70% recirculated air in the context of sufficient intake of outdoor air appeared not to increase mucosal irritation, skin or allergic reactions, or general symptoms, nor to increase the reporting of poor air quality. Thus, the findings of this study detract from the theory that change in the physical character of air due to recirculation (homogenization) causes the symptoms of the sick building syndrome.[29] Similarly, the use of a high mechanical ventilation rate in another experimental study,[4,6] which was also suggested as a cause of the homogenization of indoor air,[29] appeared not to cause these symptoms.

The present study can also be interpreted as a comparison of the effect of two levels of outdoor air supply (6 vs 20 L/s per person) on the occurrence of sick building syndrome symptoms and perceptions, if the hypothesis of the homogenizing effect of air recirculation, per se, is ignored. The role of air change as a determinant of such symptoms has been assessed earlier at least in five experimental studies.[4,6,30-33] It is evident that too small an air change allows indoor air pollutants to accumulate and that low airflows can thus be indirect determinants of both symptoms and other health problems. In a controlled trial of 46 Canadian office workers in 1980, Sterling and Sterling[30] varied the proportion of outdoor air between 87% (13% return air) and 25% (75% return air) with a constant total airflow, and observed a decrease in the occurrence of

eye symptoms related to an increase of outdoor air intake. In a controlled trial of 940 Finnish office workers carried out in 1985, a temporary reduction of originally high airflows resulted in an increase in the symptoms of the sick building syndrome[4,6] in outdoor airflows below 7.5 L/s per person. Nagda and colleagues[31] carried out an uncontrolled trial in the summer of 1989 and the winter of 1990, in which they compared the occurrence of symptoms in two levels of outdoor air supply (9 and 16 L/s per person) and reported more building-related symptoms during the low outdoor air supply than during the high outdoor air supply. In two experimental studies, no association was observed between the different levels of outdoor air supply.[32,33] Wyon[32] assessed experimentally the relation between the level of airflows and the occurrence of sick building syndrome symptoms in Malmö General Hospital staff: neither a 30% decrease nor 40% increase had an effect on the occurrence of symptoms. Menzies and colleagues[33] carried out a six-period crossover trial among 1546 Canadian office workers in 1990 "to test the hypothesis that symptoms of sick building syndrome could be reduced in mechanically ventilated office buildings by increasing supply air from 20 to 50 ft^3 per minute per person"[30(p821)] (9 and 23 L/s per person). The results were in line with the present study: the occurrence of ocular, mucosal, and general symptoms was similar with the different levels of outdoor air supply (the actual airflows were 14 and 30 L/s per person). The main difference between the two studies is that, in the Canadian study, the proportion of return air was high (68% and 93%) in both levels of outdoor air supply, whereas in the present study, 70% return air was contrasted with no return air. The role of air recirculation and outdoor air supply as determinants of the symptoms cannot be fully separated in the studies where the compared conditions include a high proportion of return air.

Sufficient air change depends essentially on human activities, sources of air pollution, net emission, and the presence of alternative exhaust routes such as operable windows and diffusion through the walls. Thus, the studies that were carried out in a small number of buildings should be used cautiously when considering the standards for minimum air change. The inconsistency in the results of different studies as to the relation between air change and the symptoms of the sick building syndrome is probably explained by differences in the indoor environment of the buildings.

Our results provided evidence that 70% recirculated air, when accompanied by an adequate intake of outdoor air, can be used without causing adverse effects. Another approach is needed to evaluate whether a ventilation system using air recirculation can be considered a risk solution in existing buildings—that is, whether the adverse health effects occur more often in buildings with such a ventilation system compared with buildings without the use of air recirculation. □

Acknowledgments

This study was supported by the Finnish Ministry of Trade and Industry and the Ministry of the Environment. Dr Jaakkola was also supported by the Yrjö Jahnsson Foundation and received a research fellowship of the Medical Research Council of the Academy of Finland.

We thank Mr Pauli Miettinen, MSc, for assistance in programming.

References

1. Sick building syndrome. *Lancet.* 1991;338:1493–1494. Editorial.
2. Burge S, Hedge A, Wilson S, et al. Sick building syndrome: a study of 4373 office workers. *Ann Occup Hyg.* 1987;31:493–504.
3. Finnegan MJ, Pickering CAC, Burge PS. The sick building syndrome: prevalence studies. *Br Med J.* 1984;289:1573–1575.
4. Jaakkola JJK. *Indoor Air in Office Building and Human Health. Experimental and Epidemiologic Study of the Effects of Mechanical Ventilation* [in Finnish; English summary]. Helsinki, Finland: University of Helsinki, Department of Public Health; 1986. Doctoral Thesis.
5. Jaakkola JJK, Heinonen OP, Seppänen O. Sick building syndrome, sensation of dryness and thermal comfort in relation to room temperature in an office building: need for individual control of temperature. *Environ Int.* 1989;15:163–168.
6. Jaakkola JJK, Heinonen OP, Seppänen O. Mechanical ventilation in office buildings and the sick building syndrome. An experimental and epidemiological study. *Indoor Air.* 1991;1:111–121.
7. Norbäck D, Rand G, Michel I, Amcoff S. The prevalence of symptoms associated with sick buildings and polluted industrial environments as compared to unexposed reference groups without expressed dissatisfaction. *Environ Int.* 1989;15:85–94.
8. Skov P, Valbjørn O, and DISG. The "sick" building syndrome in the office environment: the Danish Town Hall Study. *Environ Int.* 1987;13:339–349.
9. Skov P, Valbjørn O, Pedersen BV. Influence of personal characteristics, job-related factors and psychosocial factors on the sick building syndrome. *Scand J Work Environ Health.* 1989;15:286–292.
10. Skov P, Valbjørn O, Pedersen BV. Influence of indoor climate on the sick building syndrome. *Scand J Work Environ Health.* 1990;16:363–371.
11. Taylor PP, Dell'Acqua BJ, Baptiste MS, et al. Illness in an office building with limited fresh air access. *J Environ Health.* 1984;47:24–27.
12. Turiel I, Hollowell CD, Miksch JV, et al. The effect of reduced ventilation on indoor air quality in an office building. *Atmospheric Environ.* 1983;17:51–64.
13. *Indoor Air Pollutants: Exposure and Health Effects.* Report on a World Health Organization meeting. Copenhagen, Denmark: World Health Organization Regional Office for Europe; 1983. EURO reports and studies no 78.
14. Stolwijk JAC. The "sick building" syndrome. In: Berglund B, Berglund U, Lindvall T, Sundell J, eds. *Indoor Air:* vol 1. *Recent Advances in the Health Sciences and Technology.* Stockholm, Sweden: Swedish Council for Building Research; 1984:23–29.
15. Akimenko VV, Andersen I, Lebowitz MD, Lindvall T. The "sick" building syndrome. In: Berglund B, Berglund U, Lindvall T, Sundell J, eds. *Indoor Air:* vol 6. *Evaluations and Conclusions for Health Sciences and Technology.* Stockholm, Sweden: Swedish Council for Building Research; 1986:87–97.
16. Hedge A, Burge PS, Robertson AS, Wilson S, Harris-Bass J. Work-related illness in offices: a proposed model of the "sick building syndrome." *Environ Int.* 1989;15:143–458.
17. Mølhave L. The sick buildings and other buildings with indoor climate problems. *Environ Int.* 1989;15:65–74.
18. Mølhave L, Bach B, Pedersen OF. Human reactions during controlled exposures to low concentrations of organic gases and vapours known as normal indoor air pollutants. In: Berglund B, Lindvall T, Sundell J, eds. *Indoor Air:* vol 3. *Sensory and Hyperreactivity Reactions to Sick Buildings.* Stockholm, Sweden: Swedish Council for Building Research; 1984;431–436.
19. Norbäck D, Torgén M, Edling C. Volatile organic compounds, respirable dust, and personal factors related to prevalence and incidence of sick building syndrome in primary schools. *Br J Ind Med.* 1990;47:733–741.
20. Robertson AS, Burge PS, Hedge A, et al. Relation between passive cigarette smoke exposure and "building sickness." *Thorax.* 1988;43:263P.
21. Norbäck D, Torgén M. A longitudinal study relating carpeting with sick building syndrome. *Environ Int.* 1989;15:129–136.
22. Jaakkola JJK, Tuomaala P, Seppänen O. Textile surface materials and the sick building syndrome. *Arch Environ Health.* In press.
23. Valbjørn O, Kousgård N. Headache and mucous membrane irritation. An epidemiological study. In: Berglund B, Berglund U, Lindvall T, Sundell J, eds. *Indoor Air:* vol 2. *Radon, Passive Smoking, Particulates and Housing Epidemiology.* Stockholm, Sweden: Swedish Council for Building Research; 1984:249–254.
24. Reinikainen LM, Jaakkola JJK, Heinonen OP. The effect of air humidifica-

tion on different symptoms in office workers. An epidemiologic study. *Environ Int.* 1991;17:243–250.
25. Reinikainen LM, Jaakkola JJK, Seppänen O. The effect of air humidification on symptoms and perception of indoor air quality. A six-period cross-over trial. *Arch Environ Health.* 1992;47:8–15.
26. Lindvall T, Sundell J. Healthy buildings—conclusions and recommendations [in Swedish]. In: Dawidowicz N, Lindvall T, Sundell J, eds. *Det Sunda Huset.* Stockholm, Sweden: Byggforskningsrådet; 1987.
27. Senn S. *Cross-Over Trials in Clinical Research.* New York, NY: John Wiley & Sons; 1993.
28. Miettinen OS. *Theoretical Epidemiology. Principles of Occurrence Research in Medicine.* New York, NY: John Wiley & Sons; 1985.
29. Berglund B, Lindvall T. Sensory reactions to "sick buildings." *Environ Int.* 1986;12: 147–159.
30. Sterling E, Sterling T. The impact of different ventilation levels and fluorescent lighting types on building illness: an experimental study. *Can J Public Health.* 1983;74:385–392.
31. Nagda NL, Koontz MD, Albrecht RJ. Effect of ventilation rate in a healthy building. In: *Healthy Buildings: Proceedings of the IAQ'91 Conference.* Washington, DC: 1991;101–107.
32. Wyon D. Sick buildings and the experimental approach. *Environ Technol.* 1992; 13:313–322.
33. Menzies R, Tamblyn R, Farant J-P, Hanley J, Nunes F, Tamblyn R. The effects of varying levels of outdoor-air supply on the symptoms of sick building syndrome. *N Engl J Med.* 1993;328:821–827.

Nitrogen Dioxide Exposures inside Ice Skating Rinks

Michael Brauer, ScD, and John D. Spengler, PhD, MS

ABSTRACT

Objectives. The common operation of fuel-powered resurfacing equipment in enclosed ice skating rinks has the potential for producing high concentrations of carbon monoxide and nitrogen dioxide. Exposures to these gaseous combustion products may adversely affect the health of those inside the rink. Little information is available on pollutant concentrations under normal operating conditions.

Methods. One-week average nitrogen dioxide concentrations in 70 northeastern US rinks were measured with passive samplers during normal winter season conditions.

Results. The median nitrogen dioxide level inside rinks was 180 ppb, more than 10 times higher than the median outdoor concentration. One-week average nitrogen dioxide concentrations above 1000 ppb were measured in 10% of the rinks.

Conclusions. Considering that short-term peak concentrations were likely to have reached two to five times the measured 1-week averages, our results suggest that nitrogen dioxide levels were well above short-term air quality guidelines and constitute a public health concern of considerable magnitude. (*Am J Public Health.* 1994;84:429–433)

Introduction

In skating rinks, the operation of gasoline- or propane-powered equipment to clean and resurface the ice can lead to elevated concentrations of combustion products. Reports indicate that high concentrations of carbon monoxide in ice rinks occasionally lead to toxicity.[1–6] Recently, acute respiratory illness due to nitrogen dioxide exposure has also been reported at indoor ice rinks.[7–11]

Acute exposure to nitrogen dioxide concentrations above 5 to 10 parts per million (ppm) may produce severe cough, hemoptysis, chest pain, and pulmonary edema.[12–14] The effects of exposure to lower levels (0.1 to 1 ppm), such as those encountered in homes using gas stoves or kerosene heaters, are more debatable. Controlled exposures of healthy (nonasthmatic) individuals to nitrogen dioxide concentrations above 1000 parts per billion (ppb) (exposure for 1 hour or longer) indicate increased airway responsiveness,[15–17] whereas exposures to lower levels have not produced any effects. On the other hand, controlled chamber studies with asthmatics suggest that small changes occur in spirometric measures and airway responsiveness for short-duration exposures to 100 to 500 ppb nitrogen dioxide.[15,18,19] However, other studies have shown no respiratory effects in asthmatics following exposures to higher levels.[20,21] Consequently, asthmatics are considered to be especially susceptible to respiratory effects of nitrogen dioxide exposure.

Because controlled chamber studies have involved mostly adult subjects, epidemiological studies may have more relevance to the ice skating population since they investigated children. Neas and colleagues report an odds ratio of 1.45 for lower respiratory symptoms in children for an increase in the annual average nitrogen dioxide concentration of 15 ppb.[22] A recent meta-analysis of 11 epidemiological studies yielded similar results, suggesting a 20% increase in the odds of a lower respiratory infection for children with a prolonged increase in exposure to 16 ppb nitrogen dioxide.[23]

Although occurrences are infrequent and are typically associated with resurfacer malfunction, which produces peak nitrogen dioxide concentrations of 1000 to 3000 ppb, reports of acute nitrogen dioxide poisoning in ice rinks demonstrate that an acute exposure in this setting can lead to respiratory illness. In contrast, we sought to examine the potential public health impact of repeated exposures to nitrogen dioxide by investigating the range of concentrations encountered in skating rinks under normal operating conditions. We hypothesized that ice rinks presented the major nitrogen dioxide exposure of users and that, based on comparisons with epidemiological studies of indoor exposures, levels of nitrogen dioxide encountered in ice rinks warrant concern.

Methods

A mail-in survey was conducted in spring 1990. One hundred seven rinks

Michael Brauer is with the Department of Medicine, University of British Columbia, Vancouver, Canada. John D. Spengler is with the Department of Environmental Health, Harvard School of Public Health, Boston, Mass.

Requests for reprints should be sent to Michael Brauer, ScD, University of British Columbia, Department of Medicine: Respiratory Division, 2206 East Mall, 3rd Floor, Vancouver, BC, Canada V6T 1Z3.

This paper was accepted November 3, 1993.

TABLE 1—Measured Nitrogen Dioxide Concentrations in Enclosed Rinks

Location of Sampler	No.	Geometric Mean, ppb	Geometric SD	Range, ppb
Outdoors	54	17.6	2.4	0.44–193
Resurfacer	57	128.0	3.6	3.11–1428
Bench	57	169.0	4.0	3.11–2140
Rink sample 2	53	168.0	4.0	1.76–2470

were contacted and asked to deploy passive nitrogen dioxide sampling devices and to complete a questionnaire. Sampling tubes, instructions, and questionnaires were mailed to all participants, and follow-up contact was made by phone. Strict criteria on sample placement and handling were described: all samplers were to be deployed for the same 7-day duration at all rinks, with at least one sampler placed immediately outside the rink but away from parking lots and major traffic areas (outdoors), one sampler placed on the main resurfacer (resurfacer), one sampler placed at the scorekeeper's bench as close to breathing height as possible (bench), and one sampler placed at the opposite side of the rink from the scorekeeper's table (rink sample 2). Additionally, each sampling packet contained either a field blank or a duplicate sampler to be placed at one of the four designated sites along with the primary sampler. Nitrogen dioxide concentrations were measured with Palmes tubes diffusive samplers, which provide an integrated average of the nitrogen dioxide concentration over a defined sampling interval.[24-26] Following the sampling period, participants returned the sampling tubes to our laboratory for analysis.

Along with the sampling tubes, participants returned a completed questionnaire describing the use of the rink and resurfacing equipment, the physical characteristics of the rink, and the building's ventilation. Questions regarding rink age and seating capacity were asked for their potential use as surrogate data for building air tightness and volume, respectively. Except where indicated, the scorekeeper's bench sample was used to measure the indoor concentration of the rink. Additionally, open-air rinks were excluded from the analysis except where specifically stated. Relationships between nitrogen dioxide concentrations and rink variables obtained from questionnaires were examined by viewing scatter plots of concentration data, log-transformed concentration data, and single-factor analysis of variance of log-transformed concentration data.

Results

Valid samples and questionnaires were obtained from 70 of the 107 (65%) rinks contacted for the survey, a response rate that is in good agreement with other mail-in air pollution surveys.[27,28] Literature reports[8] and discussions with industry representatives indicate that there are more than 2000 operating ice rinks in North America. While the 70 rinks included in our survey were not intended to constitute a random representative sample, they do represent a significant proportion of operating rinks in New England.

Summary statistics of nitrogen dioxide concentrations are presented in Table 1. Samples of outdoor concentrations were substantially lower than any of the other samples for all rinks but one, with the median indoor concentration amounting to approximately 10 times the median outdoor level. Indoor concentrations from both samples in each rink were similar, indiating that nitrogen dioxide concentrations were quite homogeneous near the ice surface. Rink samples were also higher than the sample collected on the resurfacer, owing to the storage of the resurfacer either outdoors or in a garage when not in use. The resurfacer concentrations of nitrogen dioxide therefore should not be interpreted as a measure of resurfacer operator exposure.

Significant differences ($P < .05$) in nitrogen dioxide concentrations were observed between enclosed (geometric mean = 169.5 ppb; geometric standard deviation [SD] = 4.0; n = 58) and open rinks (geometric mean = 16.2 ppb; geometric SD = 1.57; n = 9), with greater differences observed if the enclosed rink did not have a ventilation system. Nitrogen dioxide concentrations in enclosed rinks without ventilation systems (geometric mean = 329.7 ppb; geometric SD = 3.7; n = 7) were higher than those in rinks with ventilation systems (geometric mean = 154.7 ppb; geometric SD = 4.0; n = 51), although the difference was not significant. The presence of a ventilation system in a rink does not necessarily mean that the system was in operation or was being operated at full capacity.

Although trends were evident, no other statistically significant differences in nitrogen dioxide concentrations were observed. Higher nitrogen dioxide concentrations were measured in rinks with propane-fueled resurfacers (geometric mean = 221.1 ppb; geometric SD = 3.3; n = 27) than in rinks using gasoline-powered resurfacers (geometric mean = 134.0 ppb; geometric SD = 4.6; n = 31); such a finding was expected because of the lower combustion temperature of propane relative to gasoline and the consequent increase in nitrogen dioxide emissions. None of the rinks tested used diesel- or electric-powered resurfacers, although these are commercially available. The presence of catalytic converters was distributed evenly between the two fuel types of resurfacers; rinks with a resurfacer containing a catalytic converter had higher nitrogen dioxide levels (geometric mean = 214.0 ppb; geometric SD = 3.3; n = 25) than rinks with a resurfacer that lacked a catalytic converter (geometric mean = 140.0 ppb; geometric SD = 4.3; n = 31). This result is explained by the oxidative properties of the catalyst, which is designed to convert carbon monoxide to carbon dioxide. Ninety-two percent of rinks used a resurfacer with a four-cylinder engine; 7% used an eight-cylinder model, and one rink used a six-cylinder model.

Rink size, as measured by seating capacity, was not a significant determinant of nitrogen dioxide concentration. This may be owing to inadequate air mixing in rinks, such that the effective volume of elevated nitrogen dioxide is smaller than the total rink volume. This is possible given the temperature stratification within enclosed rinks,[1] where cool air near the ice surface prevents effective mixing of the combustion emissions with the warmer air above. Thus, seating capacity may be a poor surrogate for rink volume. The number of resurfacings per day also was not a significant predictor of nitrogen dioxide concentration, although this was probably owing to the low sample sizes of rinks with fewer than 10 or more than 13 resurfacings per day.

Table 2 summarizes the rink characteristics.

Discussion

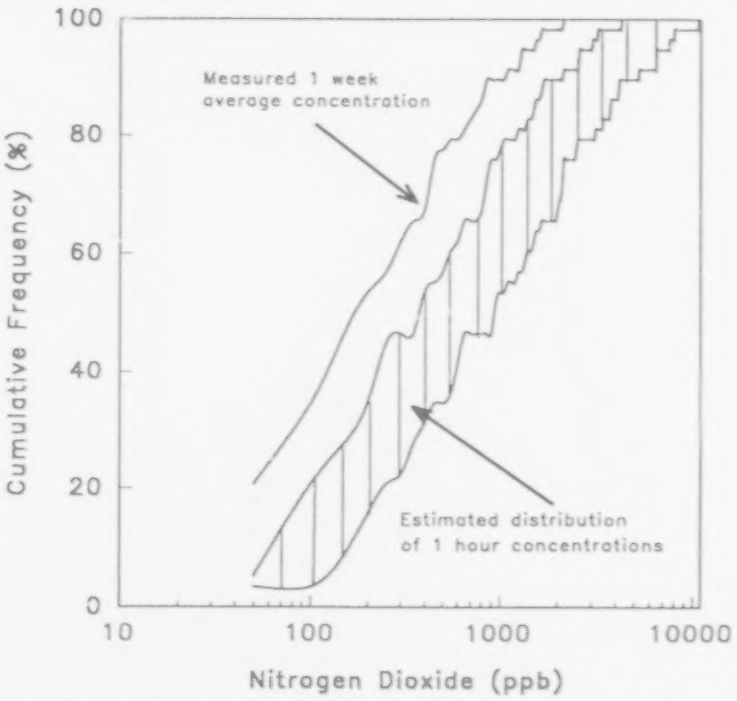
To estimate the importance of nitrogen dioxide exposures in ice rinks, we compared our measurements with those from other environments. Typical ambient nitrogen dioxide concentrations in urban areas are below 50 ppb, with infrequent hourly values above 200 ppb. Indoor concentrations are often much higher than ambient values, particularly in homes with unvented gas appliances. Since more than 40% of US homes contain gas appliances, the problem of elevated indoor nitrogen dioxide is more widespread. Extensive (2-week average) measurements in Boston homes with gas stoves[29] report median concentrations of 20 to 40 ppb, depending on the season of the measurement and the location in the home where the measurement was collected. Ninety-ninth-percentile values of approximately 100 ppb were recorded for kitchen measurements in all seasons except summer. During cooking, peak nitrogen dioxide concentrations of 450 ppb (1-hour average) are sustained for 1 to 2 hours.[30] This scenario of short-duration intermittent exposures is similar to the 1- to 3-hour exposures experienced by spectators and skaters in ice rinks; rink employees as well as coaches and figure skaters are likely to be exposed for longer periods (4 to 8 hours). Skaters are probably the most susceptible to health effects of air pollutants in an ice skating rink since ventilation rates while skating are expected to exceed normal resting breathing rates by three to five times.[4]

The detailed study by Lee and colleagues[31] adds insight into the peak nitrogen dioxide concentrations produced inside ice rinks. If the temporal distribution of nitrogen dioxide and carbon monoxide concentrations in ice rinks are similar, we estimate that 1-hour nitrogen dioxide concentrations would be two to five times the measured 1-week average (Figure 1). Assuming that the parameters assessed in our survey are typical of the more than 2000 ice rinks in operation in the United States and Canada, a rink is staffed by employees for 14 hours and visited by 300 people (for 1 to 2 hours) daily (Table 2). Therefore, 500 000 or more individuals per day may be exposed to concentrations of nitrogen dioxide that are well above the highest measured ambient and indoor levels. Repeated exposures are common, since high school and college hockey teams practice daily from September to April and recreational players skate once or twice weekly. The high levels of nitrogen dioxide encountered in ice skating rinks present an extreme exposure not likely to be encountered in other indoor, ambient, or occupational settings.

TABLE 2—Ice Rink and Resurfacer Use Characteristics, All Rinks

	Mean	SD	Range
Seating capacity	1228	1461	0–7800
Rink age, y	20	13	1–86
Rink use, h/d	14	3.4	7–20
Users per day	352	218	75–1000
Resurfacer use, min/wk	740	227	168–1260
Tuneups per year	6	8	1–52
Edger use, min/wk	104	420	5–420

Note. Numbers listed are for typical weekdays. In many rinks, weekend user numbers increase to two to five times the weekday numbers.



Note. Distribution of 1-hour concentrations is based on peak levels, which are two to five times higher than the measured 1-week average.

FIGURE 1—Cumulative frequency distribution of measured (1 week average) and estimated short-term peak (1 hour average) nitrogen dioxide concentrations in enclosed ice skating rinks.

Measurements of carbon monoxide and nitrogen dioxide collected after serious poisoning incidents indicate that levels probably exceeded even ceiling limits for occupational conditions.[32] From a regulatory standpoint, then, it is most important to eliminate incidents through corrective action and to set air quality standards for normal operating conditions at levels well below those that lead

to measurable acute effects. Minnesota has enacted air quality standards for ice arenas[33,34] stipulating that immediate corrective action must be taken when carbon monoxide exceeds 34 ppm or nitrogen dioxide exceeds 0.5 ppm (1-hour average) and requiring rinks to be evacuated when carbon monoxide exceeds 125 ppm or nitrogen dioxide exceeds 2 ppm. This standard protects against accidents where immediately dangerous levels of pollutants are present, but it is insufficient to provide a margin of safety to protect susceptible subgroups, such as asthmatics or children.

Although policy for indoor air quality has not yet been developed by the US government, short-term guidelines (1-hour average) of 200 ppb and 250 ppb nitrogen dioxide have been published for residential indoor air by the World Health Organization (WHO)[35] and the Department of National Health and Welfare Canada,[36] respectively. These guidelines are designed to protect sensitive members of the population (such as asthmatics) with an adequate margin of safety. Although these guidelines are more pertinent to ice rinks than are ambient air standards, they were developed primarily for residential environments and therefore have only limited applicability for indoor ice arenas. Moreover, because skaters are often young children, occupational standards based on a healthy adult worker are not appropriate.

Figure 1 indicates that at least 60% of the rinks included in this survey would exceed the 1-hour WHO limit of 200 ppb or the Canadian guideline of 250 ppb. In fact, it may be premature to emphasize an air quality standard for ice rinks when so many rinks are likely to be out of compliance. Instead, efforts should be directed toward education, design, and operational changes in order to mitigate the impact of resurfacers on air quality.

Strategies to maintain 1-hour nitrogen dioxide and carbon monoxide concentrations below 250 ppb and 35 ppm, respectively, are recommended. For example, warming the resurfacer engine with the exhaust vented outside will reduce indoor concentrations of carbon monoxide; however, this strategy is not effective for reducing nitrogen dioxide emissions. For many enclosed rinks, preventing short-term nitrogen dioxide levels from exceeding 250 ppb (1-hour average) will require increased ventilation and/or reduced use of resurfacing equipment. During a typical 11-minute resurfacing operation using a well-maintained resurfacer, approximately 5 L of nitrogen dioxide are released (Appendix). This volume of nitrogen dioxide would have to be diluted with 20 000 m^3 of clean air to achieve a concentration of 250 ppb. With hourly resurfacings, a typical rink building of 22 000-m^3 volume would require nearly one air exchange per hour or 13 000 cu ft of ventilation air per minute. Furthermore, mechanical ventilation systems must be designed so that the cooler air over the ice surface, which contains the highest concentrations of nitrogen dioxide, mixes adequately.

Depending on the location of the rink, additional heating of the ventilation air may be required. Across the region where most of the ice skating rinks are found (throughout Canada, the midwestern United States, and New England), heating degree-days range from 5000 to 10 000. Since indoor temperatures for ice rinks tend to be in the low 40s (°F), we used 5000 degree-days to compute the additional energy required for a rink building to achieve one air exchange per hour. For example, a gas-fueled heater with a heat exchanger at a fuel price of $0.005/cu ft would cost $6000 to $8000 annually to operate. For the approximately 50% of rinks that operate for the entire year, providing chilled dehumidified air in the summer could more than double these energy costs. Since our survey found that 10% of the enclosed rinks do not even have a ventilation system, certain rinks would also need to incur large capital costs to achieve desirable air quality.

An alternative approach is to replace existing gasoline- or propane-fueled resurfacers with electrically powered units. Since resurfacers typically have operating lifetimes of 20 to 30 years and cost $60 000 to $70 000, replacement of existing resurfacers is equivalent to an annual cost of approximately $3000.

This paper, along with reports on outbreaks of illness, points to the potential seriousness of exposures to air pollutants in ice rinks. Public health officials should recognize that inadequate ventilation and poor maintenance of the resurfacer can result in nitrogen dioxide and carbon monoxide levels that are injurious to normal healthy individuals. However, even when equipment is properly maintained, inadequate ventilation will result in nitrogen dioxide levels that can evoke airway responses in asthmatics. Health agencies are encouraged to enact requirements to prevent potentially deleterious exposures in enclosed ice skating arenas. □

Acknowledgments

The authors would like to thank the members of the New England Ice Skating Manager's Association for their participation in the survey. The help of Lauri Wast in sampler preparation and analysis is also appreciated. Kiyoung Lee provided the estimates of nitrogen dioxide emissions from resurfacers.

References

1. Anderson DE. Problems created for ice arenas by engine exhaust. *J Am Ind Hyg Assoc.* 1971;32:790–801.
2. Carbon monoxide intoxication associated with use of a gasoline-powered resurfacing machine at an ice skating rink. *MMWR.* 1984;33:49–51.
3. Carbon monoxide exposures at an ice skating rink. *MMWR.* 1986;35:435–441.
4. Levesque B, Dewailly E, Lavoie R, Prud'Homme D, Allaire S. Carbon monoxide in indoor ice skating rinks: evaluation of absorption by adult hockey players. *Am J Public Health.* 1990;80:594–598.
5. Miller RK, Ryan MC, Bilowus P. Carbon monoxide poisoning in indoor ice skating arenas. *Va Med.* 1989;116:74–76.
6. Spengler JD, Stone KR, Lilley FW. High carbon monoxide levels measured in enclosed skating rinks. *J Air Pollut Control Assoc.* 1978;28:776–779.
7. Dewailly E, Allaire S. Nitrogen dioxide poisoning at a skating rink—Quebec. *Can Dis Weekly Rep.* 1988;14:61–62.
8. Hedberg K, Hedberg CW, Iber C, et al. An outbreak of nitrogen dioxide–induced respiratory illness among ice hockey players. *JAMA.* 1989;262:3014–3017.
9. Ford DP, Rothman N, Hedberg K, et al. Nitrogen dioxide–induced respiratory illness in ice hockey players. *JAMA.* 1990; 263:3024–3025. Letter.
10. Ice hockey lung: NO_2 poisoning. *Lancet.* 1990;335:1191.
11. Smith W, Anderson T, Anderson HA. Nitrogen dioxide and carbon monoxide intoxication in an indoor ice arena—Wisconsin, 1992. *MMWR.* 1992;41:383–385.
12. National Research Council, Committee on Medical and Biological Effects of Environmental Pollutants, Subcommittee on Nitrogen Oxides. *Nitrogen Oxides.* Washington, DC: National Academy of Sciences; 1976.
13. Morrow PE. Toxicological data on NO_2: an overview. *J Toxicol Environ Health.* 1984;13:205–227.
14. Epler GR. Silo-filler's disease: a new perspective. *Mayo Clin Proc.* 1989;64:368–370.
15. Foliusbee LJ. Does No_2 exposure increase airway responsiveness? *Toxicol Ind Health.* 1992;8:273–283.
16. Frampton MW, Morrow PE, Cox C, Gibb FR, Speers DM, Utell MJ. Effects of nitrogen dioxide exposure on pulmonary function and airway reactivity in normal humans. *Am Rev Respir Dis.* 1991;143:522–527.

17. Mohsenin V. Effects of vitamin C on NO_2-induced airway hyperresponsiveness in normal subjects: a randomized double-blind experiment. *Am Rev Respir Dis.* 1987;136:1408–1411.
18. Bauer MA, Utell MJ, Morrow PE, Speers DM, Gibb FR. Inhalation of 0.3 ppm nitrogen dioxide potentiates exercise-induced bronchospasm in asthmatics. *Am Rev Respir Dis.* 1986;134:1203–1208.
19. Orehek J, Massari JP, Gayrard P, Grimaud C, Charpin J. Effect of short-time low-level nitrogen dioxide exposure on bronchial sensitivity of asthmatic patients. *J Clin Invest.* 1976;57:301–307.
20. Linn WS, Shamoo DA, Avol EL, et al. Dose-response study of asthmatic volunteers exposed to nitrogen dioxide during intermittent exercise. *Arch Environ Health.* 1986;41:292–296.
21. Bylin G, Lindvall T, Rehn T, Sundin B. Effects of short-term exposure to ambient nitrogen dioxide concentrations on human bronchial reactivity and lung function. *Eur J Respir Dis.* 1985;66:205–217.
22. Neas LM, Dockery DW, Ware JH, Spengler JD, Speizer FE, Ferris BG Jr. Association of indoor nitrogen dioxide with respiratory symptoms and pulmonary function in children. *Am J Epidemiol.* 1991;134:204–219.
23. Hasselbad V, Eddy DM, Kotchmar DJ. Synthesis of environmental evidence: nitrogen dioxide epidemiology studies. *J Air Waste Manage Assoc.* 1992;42:662–671.
24. Palmes ED. NO_2 and NO_x diffusion techniques. *Ann Am Conf Ind Hyg.* 1981;1: 263–266.
25. Yanagisawa Y, Nishimura H. A personal sampler for measurement of nitrogen dioxide in ambient air. *J Japan Soc Air Pollut.* 1980;15:316.
26. Yanagisawa Y, Nishimura H. A badge-type personal sampler for measurements of personal exposure to NO_2 and NO in ambient air. *Environ Int.* 1982;8:235–242.
27. Sexton K, Liu KS, Petreas M. HCHO concentrations inside private residences: a mail-out approach to indoor air monitoring. *J Air Pollut Control Assoc.* 1986;38: 698–703.
28. Petreas M, Liu KS, Chang BH, Hayward SB, Sexton K. A survey of nitrogen dioxide levels measured inside mobile homes. *J Air Pollut Control Assoc.* 1986;38: 647–651.
29. Ryan PB, Soczek ML, Treitman RL, Spengler JD, Billick IH. The Boston residential NO_2 characterization study, II: survey methodology and population concentration estimates. *Atmospheric Environ.* 1988;22:2115–2125.
30. Harlos DP. *Acute Exposure to Nitrogen Dioxide during Cooking or Commuting.* Boston, Mass: Harvard University, School of Public Health; 1988. PhD dissertation.
31. Lee KY, Yanagisawa Y, Spengler JD. Carbon monoxide and nitrogen dioxide levels in an indoor ice skating rink with mitigation methods. *J Air Waste Manage Assoc.* 1993;43:769–771.
32. National Institute of Occupational Safety and Health. Maximum Allowable Concentrations of Air Contaminants: Short Term Exposure Limits. 29 CFR §1910.1000 (January 1989).
33. State of Minnesota. Minnesota Rules. Enclosed Sports Arenas, pts 4635.1100–4635.2000 (1985).
34. Oatman L, Zetterlund MJ. Regulating air quality in ice arenas. *Proceedings of the Fifth International Conference on Indoor Air Quality and Climate, Toronto, Canada.* Ottawa, Ontario, Canada: Canada Mortgage and Housing Corporation; 1990;2: 373–377.
35. *Environmental Health Criteria No. 4—Oxides of Nitrogen.* Geneva, Switzerland: World Health Organization; 1977.
36. *Exposure Guidelines for Residential Indoor Air Quality.* Ottawa, Ontario, Canada: Environmental Health Directorate, Department of National Health and Welfare Canada; 1990. EHD–TR–156, Cat H49-58/1990E.

APPENDIX—Estimation of Nitrogen Dioxide Emission Rate

The average carbon monoxide emission rate as measured by Lee et al.[31] was 9 L min^{-1}. During measurements, there were 14 resurfacings lasting 10 minutes each; these resulted in a total emission of 1260 L of carbon monoxide, emitted into a rink building with a volume of 20 000 m^3, with an air exchange rate of 0.2 hr^{-1}. The resulting concentration of carbon monoxide was 13 ppm.

During this same period, 700 ppb of nitrogen dioxide were measured. Assuming that carbon monoxide and nitrogen dioxide have the same emission and concentration profile, the measured concentration of 700 ppb nitrogen dioxide corresponds to a total emission volume of 68 L nitrogen dioxide and an emission rate of approximately 0.5 L min^{-1} nitrogen dioxide. This estimated emission rate is likely to be an underestimate of the true nitrogen dioxide emission rate since this calculation assumes that carbon monoxide and nitrogen dioxide have identical decay rates when, in fact, nitrogen dioxide is likely to decay significantly faster than carbon monoxide.

The Respiratory Health Impact of a Large Urban Fire

ABSTRACT

Objectives. In July 1988, a fire destroyed a huge supermarket warehouse in Richmond, Calif, sending smoke into residential neighborhoods for nearly a week. There was no organized public health response. To evaluate the respiratory health impact on the general population, a survey of emergency room visits and hospital admissions to the two acute-care hospitals serving the population downwind was conducted.

Methods. Medical records of 489 patients meeting specified diagnostic criteria during the week of the fire and several reference periods were abstracted. Ratios of proportions for respiratory diagnoses (i.e., emergency room visits for a given diagnosis/total emergency room visits) were calculated, comparing the fire week with the reference periods, and 1988 mortality data for the area were reviewed.

Results. Ratios of proportions for emergency room visits for asthma and all lower respiratory conditions increased significantly during the fire. Respiratory-related hospitalizations also increased. However, there was no observable increase in respiratory mortality.

Conclusions. This fire was found to have had a moderate impact on the respiratory health of local residents. Public health intervention is indicated to prevent respiratory morbidity when extended exposure to structural fire smoke is predictable. (*Am J Public Health.* 1994;84: 434–438)

Michael Lipsett, MD, JD, Kirsten Waller, MD, MPH, Dennis Shusterman, MD, MPH, Susan Thollaug, RN, MPH, and Wendel Brunner, MD, PhD

Introduction

In July 1988, a fire destroyed a 580 000-sq ft supermarket distribution warehouse in Richmond, Calif. The fire began at 10:00 PM on July 11, emitting huge quantities of smoke through July 16, although it was not completely extinguished until July 19. During this time, smoke drifted north and northeast intermittently into Richmond and San Pablo, cities with a combined population of about 104 000.[1,2] Hundreds of area residents telephoned or visited hospitals and municipal offices complaining about the smoke.[3] Initially, local government agencies believed that this was a typical structural fire, and no official warnings were issued nor were other protective public health measures undertaken.

The public health impact of general population exposures to purely structural fire smoke has not been documented in the published literature. Exposure to smoke from forest fires or from a mixed wildland-suburban fire, however, has recently been reported to result in significant increases in emergency room visits for respiratory conditions.[4,5] Smoke contains various respiratory irritants, including aldehydes, particles, acid aerosols, and other compounds. Analyses of air pollution in urban areas consistently show associations between ambient particle concentrations and a variety of adverse respiratory outcomes (although the composition of smoke is undoubtedly different from routine particle sources).[6–10] Several smoke constituents, in addition, have been demonstrated to provoke bronchoconstriction.[11–13] Exposure to smoke causes acute decrements in pulmonary function and increases in airway reactivity among firefighters and fire victims.[14–16]

The purpose of this study was to estimate any excess respiratory-related hospital use associated with a major urban fire. We surveyed respiratory-related emergency room visits and hospital admissions during the fire period and two reference periods to examine epidemiologically the acute morbidity associated with the Richmond conflagration. We also compared weekly respiratory-related mortality totals for Richmond and San Pablo during and after the fire with weekly totals during the rest of 1988.

Methods

Two acute-care hospitals serving the majority of the population in the area downwind of the fire agreed to participate in this survey. One such hospital is a 42-bed facility (about 36 000 emergency room visits annually) belonging to a health maintenance organization that served 65 000 members at that time; the other is a 246-bed private hospital

Michael Lipsett and Dennis Shusterman are with the California Office of Environmental Health Hazard Assessment, Berkeley. At the time of the study, Kirsten Waller was with the Centers for Disease Control's Division of Field Epidemiology, Atlanta, Ga.; she is currently with the California Department of Health Services, Berkeley. Susan Thollaug is with the Department of Social and Administrative Health Sciences at the University of California's School of Public Health, Berkeley. Wendel Brunner is with the Public Health Division of Contra Costa County's Department of Health Services, Martinez, Calif.

Requests for reprints should be sent to Michael Lipsett, MD, JD, California Office of Environmental Health Hazard Assessment, 2151 Berkeley Way, Berkeley, CA 94704.

This paper was accepted September 30, 1993.

(about 25 000 emergency room visits annually).

The time periods selected for the survey included the fire period (the week of July 12 through 18, 1988), the week following the fire (July 19 through 25, 1988), and two reference weeks (July 8 through 14, 1986 and July 5 through 11, 1988), which were used to control for seasonal and annual trends in emergency room visits and admissions. (We were unable to use a week in 1987 because of a missing emergency room logbook.) All four periods began on a Tuesday to minimize day-of-the-week variability.

Emergency room logbooks and hospital admission listings were reviewed to identify patients who had presented with respiratory complaints or eye irritation, and to calculate total hospital admissions and emergency room visits. The medical records of patients with respiratory-related complaints were then abstracted for age, sex, date and time of emergency room visit or admission, chief complaint, and physician diagnoses. Multiple visits by the same individual were counted as separate visits.

For each emergency room visit or admission, one of the study abstractors assigned an *International Classification of Diseases*, 9th ed. (ICD-9) code to the principal physician diagnosis.[17] The diagnostic categories included the following: nasopharyngitis, sinusitis, pharyngitis, tonsillitis, laryngitis, tracheitis, upper respiratory infection (unspecified), allergic rhinitis, and upper respiratory hypersensitivity reaction (defined as upper respiratory conditions, ICD-9 codes 460–465 and 477); acute bronchitis and bronchiolitis, pneumonia, chronic obstructive pulmonary disease (including chronic bronchitis and emphysema), and asthma (defined as lower respiratory conditions, ICD–9 codes 466, 480–486, 490–492, 496, and 493); exposure to smoke and fumes from a conflagration (ICD–9 codes E891.9 and E891.2); and conjunctivitis (ICD–9 code 372). If the physician's primary diagnosis was incompatible with any of the above ICD categories, the chart was not abstracted. The diagnosis of viral syndrome was coded to one of the appropriate upper or lower respiratory ICD-9 codes, depending on the symptoms and physical findings reported; if the symptoms or findings were not referable to the respiratory tract, the chart was not abstracted. To compare the fire period with the reference periods, patients with an isolated diagnosis of smoke inhalation but with physical findings consistent with another diagnosis were coded to the latter. All emergency room visits and hospital admissions that were not coded to one of these ICD categories were defined as nonrespiratory.

We also evaluated patient charts for evidence of physical findings corroborating the physician diagnoses. We hypothesized that (owing to anxiety, legal concerns, and so forth) people may have been more likely to present with respiratory complaints but without physical findings during the fire period. This might have led to an overestimation of the respiratory effect associated with the fire. To control for this possible bias, we coded each emergency room visit or hospital admission in all time periods as "eligible" (with corroborating physical findings) or "ineligible" (without corroborating physical findings). A diagnosis of bronchitis was always coded as eligible since this diagnosis is often based on history alone. Patients given a diagnosis of smoke inhalation without any abnormal findings on physical, laboratory, or radiological exam were coded as ineligible. Although the abstractors were not blinded to the time period of the visit or admission, each case for which the determination of diagnosis or eligibility code was not straightforward was discussed in group meetings. In these discussions, only the abstractor was aware of when each visit or admission occurred, and codes were assigned based on consensus reached without regard to the time period.

Data entry and analysis were performed using Epi Info.[18] Epidemic curves were constructed by plotting daily numbers of respiratory- and nonrespiratory-related emergency room visits during the prefire, fire, and postfire periods. Ratios of proportions (RPs) of specific diagnostic categories, comparing the fire week to the two prefire reference weeks, were calculated using the following formula for each diagnostic category:

$$RP = \frac{\text{Number of visits for the diagnosis/total visits during the fire period}}{\text{Number of visits for the diagnosis/total visits during the reference periods}}$$

Chi-square tests were performed to test the homogeneity of the proportions of visits for respiratory and nonrespiratory diagnoses between (1) the two reference weeks and (2) the two hospitals.[19] The groups were found to be highly homogeneous; therefore, all results reported comprise pooled data. Confidence intervals (CIs) were computed for all relative proportions using the method of Gart and Nam.[20]

Finally, summary mortality data (obtained from the California Department of Health Services, Health Data and Statistics Branch) for Richmond, San Pablo, and El Sobrante (another small community potentially affected by the smoke) for calendar year 1988 were reviewed to ascertain whether respiratory-related deaths increased during or shortly after the fire period.

Although the smoke was visible for miles, there were unfortunately no relevant air quality data to analyze in conjunction with the health outcome information. The local air pollution control district's closest fixed-site monitor was west and upwind of the plume; the only 24-hour particulate matter sample (measured as total suspended particles) taken at that site during the fire week (on July 14) showed a reading of 52 $\mu g/m^3$, which was slightly greater than readings from samples taken 6 and 12 days earlier (46 $\mu g/m^3$ and 40 $\mu g/m^3$, respectively). No supplemental air monitoring was conducted.

Results

Hospital Survey

There were 489 charts abstracted for the four periods, representing 98.9% of all patients identified from emergency room logbooks and hospital admission records. Of these charts, 451 were for emergency room visits only, 30 were for emergency room visits resulting in hospital admission (counted as both emergency room visits and admissions in the analysis), and 8 were for direct hospital admissions. The distribution of visits and admissions by period is presented in Table 1. The numbers of daily respiratory-related emergency room visits are shown in Figure 1. (There were no differences in numbers of nonrespiratory-related emergency room visits during those periods.)

The ratios of proportions of specific diagnostic categories for emergency room visits are presented in Table 2. Emergency room patients were 1.4 times more

TABLE 1—Numbers of Emergency Room Visits and Hospital Admissions to Two Hospitals in Richmond, Calif, during a Major Urban Fire in 1988 and Several Reference Periods

	Reference Week 1[a]	Reference Week 2[b]	Fire Week[c]	Postfire Week[d]
Total respiratory-related emergency room visits[e]	87	105	160	92
Eligibles	82	100	138	NA
Ineligibles	5	5	22	NA
Nonrespiratory emergency room visits	948	1012	1009	1003
Total emergency room visits	1035	1117	1169	1095
Respiratory-related admissions[f]	9	8	15	6
Nonrespiratory admissions	153	159	149	126
Total admissions	162	167	164	132

[a]July 8–14, 1986.
[b]July 5–11, 1988.
[c]July 12–18, 1988.
[d]July 19–25, 1988 (postfire week data were used only to generate Figure 1 and not to calculate ratios of proportions).
[e]Eligible and ineligible patients combined (see explanation in text).
[f]Eligible patients only. (There were no ineligible hospitalized patients.)

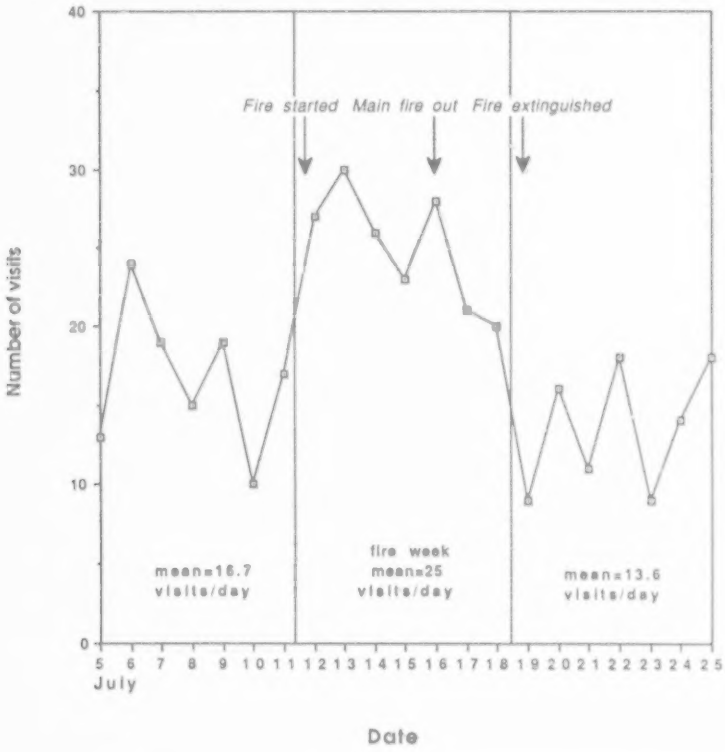


FIGURE 1—Total emergency room visits for respiratory complaints, by day, at two Richmond area hospitals, July 5 through 25, 1988.

likely to receive an eligible respiratory diagnosis during the fire period than during the reference periods. Diagnoses of lower respiratory conditions (RP = 1.55) were increased to a greater extent than those of upper respiratory conditions (RP = 1.20). Among lower respiratory conditions, diagnoses of asthma, bronchitis, and pneumonia were all elevated. The ratio of proportions for conjunctivitis was increased but was not statistically significant (data not shown). We noted no age-related differences in emergency room visits for particular respiratory diagnoses during the fire period compared with the reference periods (data not shown). The proportion of emergency room visits with respiratory diagnoses but without any corroborating physical findings (i.e., ineligible visits) increased 4.05-fold during the fire period; this increase was highly significant (95% CI = 1.92, 8.52).

For hospital admissions, the proportion with diagnoses of respiratory-related conditions was elevated during the fire period, but the confidence interval included unity (RP = 1.77; 95% CI = 0.91, 3.45). Because of the small number of hospitalized patients, we did not calculate ratios of proportions for individual diagnoses.

Community Mortality Data

In the Richmond, San Pablo, and El Sobrante areas, there were no deaths from respiratory causes in the week preceding the fire, one such death during the fire period, four during the week after the fire had been extinguished, and three the following week. All-cause mortality for these weeks totaled 16, 11, 17, and 17, respectively. During calendar year 1988, weekly totals for respiratory causes of death ranged between zero and six (mean = 2.4, SD = 1.5); and for all causes, between 11 and 39 (mean = 23.1, SD = 6.0). Visual inspection of the data indicated that there was no obvious increase in mortality during the 3 weeks after the fire started (data not shown).

Discussion

In this survey, moderate increases in the proportion of hospital visits due to respiratory conditions were found during a 1-week period when an urban population was exposed to substantial quantities of structural fire smoke. However, only the ratio of proportions for asthma was significantly elevated.

Finding an increased proportion of visits for lower respiratory conditions and for asthma in particular was not unexpected, given that smoke, chemical irritants, and airborne particles can all provoke bronchoconstriction.[11-13] In this

study, the vast majority of visits for asthma (96%) during the fire period were made by patients with a previous history of this condition. Thus, the increased ratio of proportions for asthma was not due mainly to isolated episodes of bronchospasm, although such episodes can occur in association with smoke inhalation in people without a history of asthma.[16]

Our results regarding asthma are consistent with a recent report on the impact of forest fire smoke on the general population.[4] However, in that report (which used an emergency room survey design similar to the one reported here) the authors also found significant increases in emergency room visits for upper respiratory conditions. The disparity in results may be due to the substantially greater sample size in the forest fire study, differences in exposure intensity and duration, smoke composition, demographic characteristics, or emergency room triage and diagnostic patterns. Another recent report documenting the health impact of an urban wildfire found that nearly a third of fire-related emergency room visits were for bronchospastic reactions.[5]

Numerous temporal, economic, and individual behavioral factors affect the use of emergency medical services and limit the ability of hospital-based surveys to estimate community effects of acute exposure to smoke or other pollutants. Temporal variation in hospital usage patterns could lead to the selection of inappropriate reference periods and inaccurate estimations of effect. We tried to control for these factors by selecting remote and recent reference periods that paralleled the fire period (i.e., that began on the same day of the week), and by comparing the proportion of visits due to respiratory causes rather than the absolute numbers of respiratory-related visits.

In addition, hospital-based surveys can overestimate an effect if the perception of illness severity or the motivation to seek treatment is increased during periods of exposure. This study was unique in that we tried to control for these factors by distinguishing visits that were corroborated by physical findings from visits that were not. This proved to be the single factor most strongly associated with the fire period, as the proportion of ineligible emergency room visits during that period was four times higher than that in the reference periods.

TABLE 2—Ratios of Proportions for Selected Emergency Room Diagnoses at Two Hospitals in Richmond, Calif, during a Major Urban Fire in 1988

Diagnostic Category	Ratio of Proportions (95% Confidence Interval)	Number of Visits: Fire Period	Number of Visits: Reference Periods
All respiratory conditions[a]	1.40 (1.14, 1.73)	138	182
All upper respiratory conditions	1.20 (0.86, 1.68)	53	81
Upper respiratory infections	1.14 (0.66, 1.95)	21	34
All lower respiratory conditions	1.55 (1.17, 2.05)	85	101
Asthma	1.60 (1.09, 2.35)	47	54
Bronchitis	1.64 (0.98, 2.95)	24	27
Pneumonia	1.84 (0.59, 5.69)	6	6
Ineligible visits for respiratory complaints	4.05 (1.92, 8.52)	22	10

[a]Eligible emergency room visits only (see explanation in test), fire period compared with two pooled reference periods.

Conversely, hospital-based surveys may underestimate an effect if most illnesses associated with exposure are relatively mild or if the exposed population has access to other health care services. The influence of these factors limits the sensitivity of hospital-based surveys in estimating the overall public health impact of acute exposures. Our results therefore probably represent a conservative estimate of overall respiratory morbidity in the smoke-exposed areas. In addition, some of the patients presenting during the fire period were probably not exposed to the smoke since the catchment areas of these hospitals also extend to the east and south of the plume. Thus, the case mix of respiratory complaints during the fire period would have included both exposed and unexposed individuals, resulting in a bias toward the null in the estimates of the ratios of proportions.

All-cause mortality and respiratory-related deaths in the cities of Richmond, San Pablo, and El Sobrante during the fire period and subsequent weeks were well within the usual weekly trends for calendar year 1988. Thus, the fire's impact on acute respiratory morbidity was not reflected in the area's mortality statistics. Nevertheless, smoke exposure may have contributed to the demise of individual patients.

Some, if not most, of the respiratory morbidity associated with exposure to smoke from the warehouse fire might have been prevented by appropriate interventions—for example, official advisories to persons with preexisting respiratory disease, establishment of a shelter in a part of the county unaffected by the smoke, or possibly more aggressive efforts to extinguish the fire. The Richmond–San Pablo area contains a large oil refinery and other chemical production and storage facilities, and local authorities' emergency planning has focused almost exclusively on responses to inadvertent toxic gas releases. Had the fire occurred in a chemical storage facility (as opposed to a supermarket warehouse), a more vigorous response to protect public health, presumably including evacuation, would have been triggered. In situations in which prolonged general population exposure is predictable (as was the case here), organized protective measures should be undertaken expeditiously to prevent acute respiratory morbidity, particularly among persons with preexisting disease. □

Acknowledgments

We would like to acknowledge the invaluable assistance of the medical records staffs of Kaiser and Brookside hospitals in identifying and retrieving hospital charts for review. We would also like to acknowledge Marian Gentry, RN, (Contra Costa County Department of Health Services) and the late Dr Charles Gardipee, who helped review and abstract medical records.

References

1. Richmond distribution center/Safeway tries to figure costs of fire. *San Francisco Chronicle.* July 13, 1988: A:2.
2. Smoke still a bother. *San Francisco Chronicle.* July 16, 1988: A:4.
3. Safeway a year later. *West County Times.* July 11, 1989: A:1.
4. Duclos PJ, Sanderson LL, Lipsett M. Assessment of emergency room visits for potential public health impacts of the 1987 forest fire disaster in California. *Arch Environ Health.* 1990;45:53–58.

5. Shusterman D, Kaplan JZ, Cannabarro C. Immediate health effects of an urban wildfire. *West J Med.* 1993;158:133–138.
6. Second addendum to air quality criteria for particulate matter and sulfur oxides (1982): assessment of newly available health effects information. Washington, DC: Environmental Protection Agency; 1986. EPA/600/8–86/020F.
7. Ostro BD. Air pollution and morbidity revisited: a specification test. *J Environ Econ Manage.* 1987;14:87–98.
8. Pope CA III. Respiratory hospital admissions associated with PM_{10} pollution in Utah, Salt Lake, and Cache valleys. *Arch Environ Health.* 1991;46:90–97.
9. Pope CA IV, Dockery DW, Spengler JD, Raizenne ME. Respiratory health and PM_{10} pollution. A daily time series analysis. *Am Rev Respir Dis.* 1991;144:668–674.
10. Dockery DW, Speizer FE, Stram DO, Ware JH, Spengler JD, Ferris BG Jr. Effects of inhalable particles on respiratory health of children. *Am Rev Respir Dis.* 1989;139:587–594.
11. Widdecombe JG, Kent OC, Nadel JA. Mechanism of bronchoconstriction during inhalation of dust. *J Appl Physiol.* 1962;17:613–616.
12. Nadel JA, Comroe JH Jr. Acute effects of inhalation of cigarette smoke on airway conductance. *J Appl Physiol.* 1961;16:713–716.
13. Sheppard D. Chemical agents. In: Murray JF, Nadel JA, eds. *Textbook of Respiratory Medicine.* Philadelphia, Pa: WB Saunders Co; 1988:1631–1645.
14. Sherman CB, Barnhart S, Miller MF, et al. Firefighting acutely increases airway responsiveness. *Am Rev Respir Dis.* 1989; 140:185–190.
15. Sheppard D, Distefano S, Morse L, Becker C. Acute effects of routine firefighting on lung function. *Am J Ind Med.* 1986;9:333–340.
16. Kinsella J, Carter R, Reid WH, Campbell D, Clark CJ. Increased airways reactivity after smoke inhalation. *Lancet.* 1991;337: 595–597.
17. US Department of Health and Human Services. *The International Classification of Diseases.* 9th rev. *Clinical Modification.* Rockville, Md: US Public Health Service, Health Care Financing Administration; 1980. DHHS publication PHS 80–1260.
18. Dean AG, Dean JA, Burton AH, Dicker RC. *Epi Info, Version 5: A Word Processing, Database, and Statistics Program for Epidemiology on Microcomputers.* Stone Mountain, Ga: USD Inc; 1990.
19. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley & Sons; 1981:160–165.
20. Gart JJ, Nam J. Approximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness. *Biometrics.* 1988;44:323–338.

Dioxins and Dibenzofurans in Adipose Tissue of the General US Population and Selected Subpopulations

John E. Orban, PhD, John S. Stanley, PhD, John G. Schwemberger, MS, and Janet C. Remmers, MS

ABSTRACT

Objectives. The Environmental Protection Agency's National Human Adipose Tissue Survey (NHATS) was conducted in fiscal year (FY) 1987 to (1) estimate average concentrations of polychlorinated dibenzo-*p*-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs) in the adipose tissue of humans in the US population, (2) identify differences in average concentrations among subpopulations, and (3) compare average concentrations with those from the FY 1982 NHATS.

Methods. Population estimates of the average levels of PCDDs and PCDFs were established on the basis of 865 human adipose tissue specimens collected in FY 1987. Average levels among subpopulations were compared.

Results. The average concentration of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in the adipose tissue of the US population was 5.38 pg/g, increasing from 1.98 pg/g in children under 14 years of age to 9.40 pg/g in adults over 45. The effect of age was significant for nine compounds. Regional differences in the levels of 2,3,4,7,8-pentachlorinated dibenzofurans were statistically significant, but there were no significant differences associated with sex or race.

Conclusions. The survey provides a baseline of average levels of PCDDs and PCDFs in the adipose tissue of humans in the US population. (*Am J Public Health.* 1994;84: 439–445)

Introduction

The potential for human exposure to polychlorinated dibenzo-*p*-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs) has been extensively studied over the past several years. The PCDDs and PCDFs are recognized as ubiquitous contaminants in many areas of North America and have been determined to originate from numerous sources, including the emissions from incineration sources (municipal, hospital, and hazardous waste incinerators), effluents and solid wastes from pulp and paper processing, metal reclamation, petroleum refining, wood preservation, and the production of specific chlorinated chemicals. The PCDDs and PCDFs are stable lipophilic compounds and, as a result, persist in the environment and bioaccumulate. Human exposure can occur through a number of pathways via inhalation, ingestion, or dermal contact.

The presence of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in human tissues was first reported as a result of investigations of human exposures to the herbicide Agent Orange.[1–9] As information was developed on the presence of PCDDs as by-products from other processes, a number of studies were initiated to assess the potential for human exposure to these compounds. These studies have included the evaluation of Vietnam veterans exposed to Agent Orange[1–9]; residents of Seveso, Italy, following the explosion of a trichlorophenol production plant[10,11]; residents of Times Beach, Mo, exposed to contaminated oil spread on roads and common areas[12–15]; individuals exposed to polychlorinated biphenyl fires[16]; of consumers of contaminated food products[17]; and workers from chemical production facilities.[18–20] Most of these studies focused on comparing levels of PCDDs and PCDFs in tissues from both exposed and nonexposed populations; a few focused on general populations, but these studies were limited in size and scope.[21–27]

One unique monitoring program for assessing human exposure to potentially toxic compounds is the National Human Adipose Tissue Survey (NHATS), operated since the early 1970s by the US Environmental Protection Agency (EPA). Human adipose tissues are obtained through a network of pathologists and medical examiners to provide a representative sample of the general US population. Preliminary data on the levels of PCDDs and PCDFs in this population were previously generated using NHATS tissue samples collected in fiscal year (FY) 1982.[28–30]

The NHATS specimens from the FY 1987 collection have also been analyzed for PCDDs and PCDFs, and a study design similar to that employed in FY 1982 was used. This paper describes the objectives, sampling design, and analytical efforts required to achieve the general US population estimates, and it presents the estimated average concentrations of specific PCDD and PCDF isomers for the various demographic

John E. Orban is with Battelle Memorial Institute, Columbus, Ohio. John S. Stanley is with Midwest Research Institute, Kansas City, MO. John G. Schwemberger and Janet C. Remmers are with the Office of Pollution Prevention and Toxics, US Environmental Protection Agency, Washington, DC.

Requests for reprints should be sent to Janet C. Remmers, MS, US Environmental Protection Agency, Mail Code 7404, 401 M St, SW, Washington, DC 20460.

This paper was accepted July 26, 1993.

TABLE 1—Numbers and Percentages of Specimens Collected and Composite Samples Analyzed in the FY 1987 National Human Adipose Tissue Survey, by Categories

	Specimens Collected		Composites Analyzed		1980 Census Population
	No. (n = 865)	%	No. (n = 48)	%	%
Census region					
North Central	175	20	15	31	26
Northeast	296	34	9	19	22
South	289	33	16	33	33
West	105	12	8	17	19
Age group					
0–14 y	146	17	11	23	23
15–44 y	318	37	17	35	46
45+ y	401	46	20	42	31
Sex					
Male	436	50	16	52[a]	49
Mixed[b]	...	...	17	...	...
Female	429	50	15	48	51
Race group					
Caucasian	707	82	8	100	83
Mixed[b]	...	...	40	...	...
Non-Caucasian	158	18	0	0	17

[a]The percentage of composites in each sex and racial group is calculated from the total number of pure composites. For example, 16 of the 31 (52%) pure sex composites contain only specimens from male donors.

[b]A mixed composite contains specimens from both male and female donors or both Caucasian and non-Caucasian donors.

populations studied. Further details on the FY 1987 NHATS are available in an EPA final report.[31]

The specific objectives of the FY 1987 NHATS study were as follows:

- To estimate the average concentrations of PCDDs and PCDFs in the adipose tissue of humans in the general US population and various demographic subpopulations.
- To determine if there are significant differences among the average concentrations of PCDDs and PCDFs in the adipose tissue of individuals from different demographic subpopulations as defined by the individuals' geographic region, age, race, and sex.
- To compare the estimated average concentration levels of PCDDs and PCDFs found in the FY 1987 NHATS with estimates from the FY 1982 NHATS.

Methods

Study Design

The FY 1987 NHATS used a multistage sampling plan to collect human adipose tissue specimens from October 1986 through September 1987. Because of the invasive nature of the process required to collect the adipose specimens from living persons, the sampling population was limited to cadavers and surgical patients.

The 48 conterminous states were stratified into 17 geographical areas defined by the intersection of the 9 US census divisions and the 10 EPA regions. Although the sampling effort was divided among 17 geographic strata, the strata were later combined into the four US census regions (Northeast, North Central, South, and West) to simplify the presentation of results.

A sample of metropolitan statistical areas was selected with probabilities proportional to each area's population. Forty-seven such areas were selected from 14 of the 17 sampling strata for the FY 1987 NHATS; areas were not selected from the other 3 sampling strata (Arizona-Nevada, Idaho, and New Mexico) because of their small population size.

One or more cooperators (hospital pathologists or medical examiners) from each metropolitan statistical area supplied tissue specimens using quotas based on categories defined by the donor's age group (0 to 14 years, 15 to 44 years, and 45+ years), race (Caucasian and non-Caucasian), and sex. The subquotas are proportional to the 1980 US census population counts for each sampling stratum. The donors and tissue specimens were selected in a nonprobabilistic manner by the cooperators.

The guidelines for collecting specimens from cadavers specify that the death should have been caused by sudden traumatic injury such as cardiac arrest, car accident, or gunshot wound. Also, institutionalized individuals, persons known to be occupationally exposed to toxic chemicals, persons who died of pesticide poisoning, and persons suffering from cachexia were excluded from NHATS.

Interpretation of the statistical estimates derived from the NHATS data depend on several assumptions: (1) that the concentrations of toxic substances in the adipose tissue of cadavers and surgical patients are comparable to those in the general population; (2) that no significant bias is introduced by selecting only urban hospitals and medical examiners (i.e., those located in metropolitan statistical areas); and (3) that no systematic bias is introduced by the fact that the cooperators are not randomly selected and the donors and specimens are nonprobabilistically sampled according to prespecified quotas.

Of 956 specimens collected by the cooperators in FY 1987, 865 were designated for chemical analysis. Some specimens were excluded from the analysis because several cooperators exceeded their quotas for specimens in specific age-race-sex categories.

Prior to chemical analysis, each of the 865 "design" specimens were assigned to 48 composite samples. The total number of composite samples was limited by the budget for chemical analysis. Using composite samples helps to minimize the random effects of sampling error because the concentration of a target analyte in the composite sample represents the average concentration for individuals contributing specimens. Compositing criteria were established to achieve the study objectives of estimating and comparing average concentrations in selected subpopulations. These criteria specified that composites should be created using only specimens from donors in the same age group and the same US Census Division; however, composites could also be created using specimens from donors of different sexes or race groups.

Table 1 presents the numbers of specimens collected and of composite

samples analyzed within each category of the four analysis factors (geographic region, age, race, and sex). The percentages of total specimens collected and of total composites analyzed are also presented, along with the corresponding population percentages from the 1980 US Census. The statistical treatment of the data included adjustments to account for the differences between the population percentages and the percentages of specimens collected and composites analyzed.

In accordance with the abovementioned criteria, each composite sample is composed entirely of specimens from donors of a specific geographic region and age group. This ensures maximum precision for estimating differences in body burden levels among populations from different geographic regions and age groups.

The percentages of specimens from different donor sex and racial groups in the composite samples vary across the design depending on the actual collection of specimens within specific demographic subpopulations. In particular, it was not always possible to create composite samples containing only specimens from the same sex or race group. Seventeen of the 48 composites contained specimens from both sexes and 40 contained specimens from both race groups. However, it was still possible to estimate average concentrations for different sexes and racial groups by using the percentages of specimens from each demographic subpopulation as covariates in the statistical analysis.

Sample Preparation and Chemical Analysis

The 48 composite samples were prepared in the analysis laboratory to determine low picogram-per-gram (parts per trillion, ppt) levels of PCDDs and PCDFs using high-resolution gas chromatography/high-resolution mass spectrometry. Aliquots of individual specimens were combined according to the compositing design derived from statistical requirements of the study. Each of the resultant 48 composite samples was assigned a unique identification number and then randomly assigned to one of five sample analysis batches, along with the quality control samples (spiked and unspiked) that were prepared from human adipose tissue.

The sample preparation and chemical analysis procedures are described in detail in the EPA's final report.[31] The procedures were used to determine the tetrachloro through octachloro analogs of 2,3,7,8-substituted PCDDs and PCDFs. The quality control specimens incorporated with the sample batches allowed the method performance to be thoroughly assessed. Details of the quality control procedures and method performance results have been previously reported.[27,31]

Analysis

The statistical model applied to the composite sample data assumes that the four analysis factors (census region, age, sex, and race) have fixed additive effects on the average PCDD and PCDF concentrations in the specimens. This assumption creates 48 subpopulations defined by the $4 \times 3 \times 2 \times 2$ unique binations of categories. Because the composite sample concentration represents the average concentration of the contributing specimens, this assumption also results in a model for composite samples in which the measured concentration is also a linear function of the parameters associated with the effects of each of the four analysis factors.

There are also three ancillary factors that have random effects on NHATS data: (1) sampling of metropolitan statistical areas, (2) sampling of individuals within metropolitan statistical areas, and (3) measurement of analyte concentrations in the composite samples. The first two ancillary factors have random effects on the actual concentrations in individual specimens. These include the effects of selecting donors as well as of selecting specimens from individual donors. The third factor has a random effect on the measured concentrations of composites.

The method of iteratively weighted generalized least squares was used to estimate the model parameters,[32] which in turn were used to calculate predicted concentrations for each of the 48 subpopulations. Finally, marginal population averages for the different analysis categories were estimated by taking weighted averages of the appropriate predicted concentrations. Standard errors were calculated for each of the marginal averages.

Hypothesis tests were performed using the likelihood ratio method to determine if average concentrations differ significantly by any of the geographic or demographic factors. For example, the hypothesis that average concentration levels do not differ among the four census regions was tested against the alternative hypothesis that there is at least one pair of regions for which the average concentrations are different. Similar tests were performed for the effects of age, sex, and race.

The likelihood ratio method assumes that sampling and measurement errors are normally distributed. The additive effect of compositing specimens suggests that this assumption is justified because specimen sampling errors are averaged in the composite sample. Analysis of the residuals from the fitted model confirmed that the normal distribution of errors was a reasonable choice.

Results

The target compounds for the FY 1987 NHATS included 7 PCDDs and 10 PCDFs. However, statistical analyses of population averages and demographic effects were performed only on compounds that met two criteria: (1) their chemical analyses had to satisfy preestablished data quality objectives for at least 30 of the 48 composite samples, and (2) they had to be detected in at least 50% of the composite samples meeting those objectives. The data quality objectives were based on expected ion ratios (±20%), absence of interferences, and acceptable internal standard recovery (40% to 150%). Only data that met all the objectives were included in the statistical analysis.

For each of the 17 target compounds, Table 2 presents the number of composites meeting the data quality objectives, the average detection limit, and the percentage of composites in which the compound was detected. Six PCDDs (1,2,3,4,7,8-hexachlorodibenzo-*p*-dioxin [HxCDD] and 1,2,3,6,7,8-HxCDD are combined because they could not be separated in the chemical analysis) and three PCDFs met the criteria for conducting the statistical analyses. Each of these nine compounds met the data quality objectives for at least 32 composites and were detected in at least 92% of the composites. Also presented in Table 2 are the minimum, median, and maximum measured concentrations of the 17 target compounds found in the composite samples. The concentrations in the composite samples represent average concentrations of approximately 18 individual specimens used to create each composite. Therefore, the range of concentrations among individuals in the general population is expected

TABLE 2—Summary of Data for FY 1987 National Human Adipose Tissue Survey Composite Samples

Analyte	Composite Samples Meeting All DQOs, n	Average LOD (pg/g)	Composites in Which Compound Was Detected, %[a]	Statistical Analysis[b]	Concentration Range (pg/g)[c]		
					Minimum	Median	Maximum
2,3,7,8-TCDF	33	0.205	100	Yes	0.893	1.89	3.88
2,3,7,8-TCDD	36	0.291	97	Yes	<0.980[d]	6.54	15.1
1,2,3,7,8-PeCDF	43	0.434	14	No	<0.066[d]	0.249	1.42
2,3,4,7,8-PeCDF	39	0.385	95	Yes	<0.264[d]	9.21	29.2
1,2,3,7,8-PeCDD	35	1.34	97	Yes	<2.44[d]	10.2	24.4
1,2,3,4,7,8-HxCDF	9	0.706	89	No	<3.11[d]	7.30	17.0
1,2,3,6,7,8-HxCDF	37	0.741	92	Yes	<0.556[d]	5.03	14.3
2,3,4,6,7,8-HxCDF	17	1.00	18	No	<0.337[d]	0.479	2.49
1,2,3,7,8,9-HxCDF	45	0.877	2	No	<0.290[d]	0.341	1.98
1,2,3,4,7,8/ 1,2,3,6,7,8-HxCDD	41	1.25	100	Yes	13.3	76.1	174.0
1,2,3,7,8,9-HxCDD	39	1.23	97	Yes	<3.86[c]	11.5	22.0
1,2,3,4,6,7,8-HpCDF	27	1.00	89	No	<1.15[d]	17.7	32.5
1,2,3,4,7,8,9-HpCDF	46	1.41	4	No	<0.731[d]	0.715	1.74
1,2,3,4,6,7,8-HpCDD	42	1.07	100	Yes	20.9	110.0	230.0
OCDF	23	1.67	30	No	<0.680[d]	1.19	13.2
OCDD	32	3.75	100	Yes	152.0	838.0	1630.0

Note. DQO = data quality objective; LOD = average detection limit; TCDF = tetrachlorodibenzo-*p*-furan; TCDD = tetrachlorodibenzo-*p*-dioxin; PeCDF = pentachlorinated dibenzofuran; PeCDD = pentachlorodibenzo-*p*-dioxin; HxCDF = hexachlorodibenzofuran; HxCDD = hexachlorodibenzo-*p*-dioxin; HpCDF = heptachlorodibenzofuran; HpCDD = heptachlorodibenzo-*p*-dioxin; OCDF = octachlorodibenzofuran; OCDD = octachlorodibenzo-*p*-dioxin.

[a] Percentage of composite samples meeting all DQOs.

[b] Statistical analysis performed if at least 30 samples met all DQOs and percentage detected exceeds 50%.

[c] Not detected (ND) concentrations are replaced by LOD/2.

[d] The minimum concentration is less than the minimum reported LOD.

to be much greater than that observed for the composite samples.

Population Estimates

Table 3 presents the estimated average concentrations of the six PCDDs and three PCDFs noted above that met the criteria for analysis. Averages are provided for the entire US population as well as for the categories defined by the four analysis factors. Also presented is the relative standard error (percentage) for each estimate. The relative standard errors for the estimates of national average concentrations were 6% or less for the dioxins and 13% or less for the furans. All estimates were adjusted for population percentages based on the 1980 US census.

The average concentration of 2,3,7,8-TCDD in the US population is estimated to be 5.38 pg/g (±6%). The estimates for various subpopulation averages are presented even if the effects of the factors were not found to be statistically significant. For example, the regional estimates of 2,3,7,8-TCDD average concentrations range from 4.54 pg/g in the West to 6.02 pg/g in the Northeast. However, the differences among regions for this analyte were not found to be statistically significant.

The age group estimates in Table 3 suggest that the concentrations of nearly all analytes increase with the age of the donor. Notice that the highest concentration of each analyte occurs in the oldest age group. In fact, the hypothesis tests confirmed that there are statistically significant differences among age groups for all nine analytes.

Also presented in Table 3 are the International Toxicity Equivalents (I-TEQs) calculated according to the International Toxicity Equivalency Factor (I-TEF) method.[33] This method, adopted by EPA, is a biologically based procedure for estimating risks associated with exposures to complex mixtures of PCDDs and PCDFs. The I-TEQ is calculated as the sum of the I-TEF values times the estimated concentrations of the specific PCDDs and PCDFs. The isomers in Table 2 are assigned I-TEF values ranging from 0.001 to 1.0. The largest values are assigned to 2,3,7,8-TCDD (I-TEF = 1.0) and to 1,2,3,7,8-pentachlorodibenzo-*p*-dioxin (PeCDD) and 2,3,4,7,8-pentachlorinated dibenzofuran (PeCDF) (I-TEF = 0.5). All other isomers have I-TEF values of 0.1 or less.

In Table 3 the I-TEQs were calculated for the entire US population and for the various subpopulations. The calculations used the predicted average concentrations within each subpopulation for the nine isomers analyzed statistically. Because there were insufficient data on the remaining isomers to estimate average concentrations within each subpopulation, the simple average composite concentrations were used in the I-TEQ calculation. The isomers that were not statistically analyzed accounted for approximately 4% (I-TEQ = 1.1) of the total I-TEQs.

Because the statistical uncertainties of the I-TEFs have not been determined, Table 3 does not list the relative standard errors for the I-TEQs. However, they are expected to be slightly higher than those listed for 2,3,7,8-TCDD.

Analysis of Demographic Factors

Statistical hypothesis tests were performed on each of the target analytes to determine if there are statistically significant differences in average concentrations between individuals from different geographic regions, age groups, sex groups, and race groups. Table 4 lists the attained significance level—the smallest significance level that will refute the hypothesis that there are no differences between population averages—from the statistical tests associated with the four

TABLE 3—Estimated Average Concentrations (pg/g) with Relative Standard Errors (RSEs, %) for Selected PCDDs and PCDFs from FY 1987 National Human Adipose Tissue Survey Composite Samples

Analyte	Entire Nation 100[a]	Census Region NE 22	NC 26	S 33	W 19	Age 0–14 23	15–44 46	45 + 31	Race Caucasian 83	Non-Caucasian 17	Sex Male 49	Female 51
PCDDs												
2,3,7,8-TCDD, pg/g	5.38	6.02	5.72	5.19	4.54	1.98	4.37	9.40	4.62	9.15	4.22	6.48
(RSE, %)	(6)	(10)	(11)	(11)	(16)	(41)	(12)	(4)	(14)	(29)	(11)	(8)
1,2,3,7,8-PeCDD, pg/g	10.7	11.3	10.8	10.4	10.3	3.30	9.33	18.2	10.6	11.4	10.8	10.6
(RSE, %)	(4)	(8)	(7)	(8)	(12)	(22)	(7)	(4)	(7)	(23)	(6)	(7)
1,2,3,4,7,8/ 1,2,3,6,7,8-HxCDD, pg/g	75.1	75.5	82.8	66.9	78.7	23.4	70.9	120	71.7	92.0	73.7	76.5
(RSE, %)	(4)	(8)	(6)	(7)	(8)	(23)	(6)	(3)	(7)	(19)	(6)	(5)
1,2,3,7,8,9-HxCDD, pg/g	11.7	10.7	12.7	11.5	11.7	6.13	10.8	17.1	11.5	12.3	11.0	12.3
(RSE, %)	(4)	(9)	(7)	(7)	(10)	(18)	(7)	(4)	(8)	(28)	(7)	(6)
1,2,3,4,6,7,8-HpCDD, pg/g	110	120	115	105	103	45.7	99.8	174	108	123	104	117
(RSE, %)	(3)	(7)	(6)	(6)	(9)	(11)	(5)	(3)	(5)	(14)	(5)	(4)
1,2,3,4,6,7,8,9-OCDD, pg/g	724	765	759	738	608	215	692	1150	718	755	676	771
(RSE, %)	(4)	(7)	(6)	(6)	(15)	(17)	(7)	(5)	(5)	(15)	(7)	(5)
PCDFs												
2,3,7,8-TCDF, pg/g	1.88	1.85	2.00	1.53	2.33	1.97	1.45	2.45	1.81	2.21	1.84	1.91
(RSE, %)	(7)	(14)	(14)	(15)	(14)	(11)	(15)	(7)	(12)	(33)	(11)	(11)
2,3,4,7,8-PeCDF, pg/g	9.70	13.7	10.5	9.51	4.49	1.87	8.00	18.0	9.25	11.9	8.70	10.7
(RSE, %)	(8)	(13)	(13)	(15)	(47)	(100)	(15)	(6)	(16)	(44)	(13)	(12)
1,2,3,6,7,8-HxCDF, pg/g	5.78	7.15	5.57	5.80	4.49	1.80	4.59	10.5	5.60	6.68	5.00	6.52
(RSE, %)	(13)	(27)	(24)	(25)	(43)	(83)	(26)	(13)	(25)	(73)	(24)	(19)
I-TEQ[b]	27.9	31.1	29.7	26.6	24.4	9.68	24.6	46.5	26.5	35.2	26.1	29.9

Note. For list of abbreviations, see note to Table 2.
[a]Population percentage based on 1980 US census.
[b]International Toxicity Equivalent (I-TEQ) includes an adjustment of 1.1 based on average composite concentrations of seven additional PCDFs.

analysis factors. For example, the differences among estimated averages of 2,3,7,8-TCDF in the four census regions could only be considered significant at the .15 level of significance. On the other hand, the differences in age group averages are significant at the .002 level. A 5% (.05) level of significance is generally the level used to declare statistical significance.

It is clear from Tables 3 and 4 that there are statistically significant differences among the average concentrations in the three age groups. As previously noted, the highest concentration of each analyte is found in the oldest age group (45+ years) and, except for TCDF, the lowest is found in the youngest age group (0 to 14 years). The only other statistically significant finding, at the .05 level of significance, was that there are regional differences in the average concentrations of 2,3,4,7,8-PeCDF. The average concentration in the western census region was estimated to be 4.49 pg/g compared with the national average of 9.70 pg/g. The highest average concentrations of 2,3,4,7,8-PeCDF were found in the Northeast (13.7 pg/g).

The differences in the estimated average concentrations between Caucasian and non-Caucasian and between male and female donors were not statistically significant for any of the modeled compounds.

Comparison of FY 1982 and 1987 NHATS Studies

One of the objectives of the FY 1987 NHATS study was to compare the levels of PCDDs and PCDFs with the baseline levels established in the FY 1982 NHATS study. The earlier NHATS results were based on the analysis of 46 composite samples created from 763 specimens. The study designs of the two studies were similar except there were 47 metropolitan statistical areas in the FY 1987 design compared with 35 in FY 1982; however, 30 of those 35 areas were included in the later survey.

The major differences between the FY 1982 and 1987 studies were the result of advances in the analytical methodology. The most significant change involved the use of additional internal quantitation standards. There were two internal standards used in FY 1982 and nine in FY 1987. As a result, the compounds that are most directly comparable between the two studies are 2,3,7,8-TCDD, and 2,3,7,8-octachlorodibenzo-p-dioxin (OCDD).

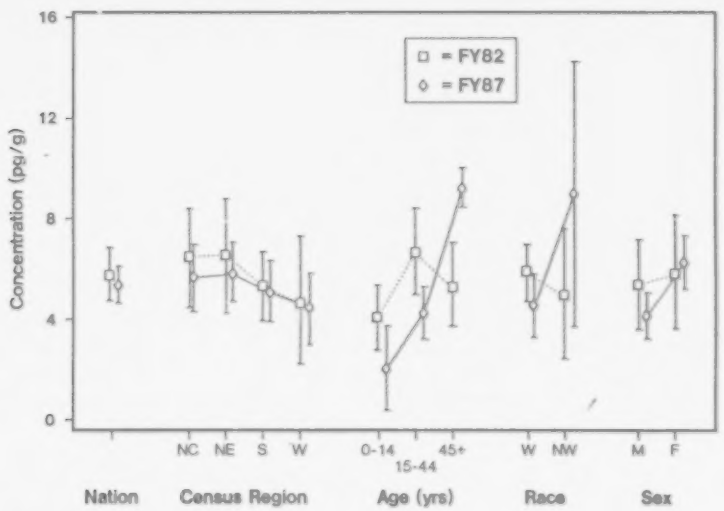
Figure 1 is a graphical representation of the demographic profiles of 2,3,7,8-TCDD from the FY 1982 and 1987 NHATS studies. The estimated average concentrations within each demographic subpopulation are connected with straight lines. The vertical lines define approximate 95% confidence intervals of the estimates. Estimated national averages with 95% confidence intervals are also plotted.

The statistical analysis reveals that there was no significant difference between the national average concentrations of 2,3,7,8-TCDD in human adipose tissue determined in the FY 1982 and 1987 NHATS studies. Also, there were no significant differences in the profiles with respect to region, sex, and race. On the other hand, only the FY 1987 NHATS demonstrated that the levels of 2,3,7,8-TCDD increase consistently with

TABLE 4—Significance Levels from Hypothesis Tests for Differences between Demographic Groups for National Human Adipose Tissue Survey FY 1987 PCDDs and PCDFs

Analyte	Analysis Factors			
	Census Region	Age	Sex	Race
2,3,7,8-TCDF	.150	.002***	.918	.672
2,3,7,8-TCDD	.247	<.001***	.232	.341
2,3,4,7,8-PeCDF	.011**	<.001***	.603	.605
1,2,3,7,8-PeCDD	.960	<.001***	.627	.593
1,2,3,6,7,8-HxCDF	.847	<.001***	.648	.795
1,2,3,4,7,8/ 1,2,3,6,7,8-HxCDD	.087*	<.001***	.934	.395
1,2,3,7,8,9-HxCDD	.289	<.001***	.597	.725
1,2,3,4,6,7,8-HpCDD	.299	<.001***	.424	.393
OCDD	.655	<.001***	.495	.520

Note. For list of abbreviations, see note to Table 2.
*Indicates significance at the .10 level.
**Indicates significance at the .05 level.
***Indicates significance at the .01 level.



Note. W = White; NW = non-White.

FIGURE 1—Profile plots of average 2,3,7,8-tetrachlorodibenzo-p-dioxin concentrations with 95% confidence intervals for FY 1982 and FY 1987 adipose tissue samples.

the age of the donors. The average concentration in the later NHATS increases from 1.98 pg/g in the youngest group to 9.4 pg/g in the oldest. Smaller differences exist between the age group averages from the earlier NHATS.

Similar results were obtained in comparing OCDD levels between FY 1982 and FY 1987. The estimated national average concentration of OCDD in FY 1982 was 768 pg/g with a standard error of 79.7. The FY 1987 national estimate was 724 pg/g with a standard error of 28.6.

Discussion

The FY 1987 NHATS established nationwide and regional estimates of average PCDD and PCDF concentrations in human adipose tissue. Although the accuracy of these estimates depends on the extent to which cadavers and surgical patients represent the general population, NHATS is the most comprehensive study of PCDD and PCDF exposures in the general population performed to date.

The national estimate of 5.38 pg/g for 2,3,7,8-TCDD in human adipose tissue obtained in NHATS compares favorably with estimates derived in recent studies[12,14,15,21,34] that focus on regional populations. The studies report average concentrations ranging from 4.6 pg/g in California[25] to 8.2 pg/g in New Jersey and Missouri.[34] However, most of the studies did not use specimens from donors that were representative of the general population in terms of age; in particular, most studies did not include specimens from donors under 14 years of age. This, along with the differences in sample collection periods, may explain the slightly higher reported concentrations compared with those in the FY 1987 NHATS.

The FY 1987 NHATS demonstrated that the average body burden levels of PCDDs and PCDFs in the general population increase with age. For example, the average concentration of 2,3,7,8-TCDD in adipose tissue was found to increase at an average rate of 0.83 pg/g per decade for individuals under age 31 and at an average rate of 1.52 pg/g per decade for the older population. Other studies[12,15,21] have estimated the rate of increase to be approximately 2.0 pg/g per decade within older populations.

While age was found to be a significant factor in predicting the average levels of PCDDs and PCDFs in adipose tissue, there was no statistical evidence of differences in the average levels for populations representing different sexes and racial groups nationwide. However, the NHATS could not adequately examine the effects of these factors on a regional basis. Studies in California[25] and Missouri[12] have found statistically significant differences in the average levels of 2,3,7,8-TCDD among men and women.

The human health risk of exposures to PCDDs and PCDFs is of great concern, but it is a topic of much debate.[35] A recent cohort study of exposed workers[36] suggests the possible association between exposure to 2,3,7,8-TCDD and increased mortality from soft tissue sarcoma and cancers of the respiratory system. Although the results are inconclusive, it is clear that future research must involve the determination

of the exposures to PCDDs and PCDFs as well as the incidence of adverse health effects. The FY 1987 NHATS provides important baseline results that will be needed for this research. □

Acknowledgments

We are grateful to Ms Cindy Stroup and Dr Joseph Breen of the EPA's Office of Pollution Prevention and Toxics for their guidance and support for this work. Also we recognize the technical contributions of Dr Robert Lordo, Dr Ron Menton, Ms Tamara Collins, Ms Pam Hartford, and Ms Claire Matthews of Battelle Memorial Institute; Ms Karin Bauer, Mr Paul Cramer, and Dr Jerry Flora of Midwest Research Institute; Ms Barbara Leczynski of EPA's Office of Pollution Prevention and Toxics; Dr Al Unger of SAIC; and Mr John Rogers of Westat.

References

1. Centers for Disease Control Veterans Health Studies Group. Serum 2,3,7,8-tetrachloro-dibenzo-p-dioxin levels in US Army Vietnam-era veterans. *JAMA.* 1988; 260:1249–1254.
2. Centers for Disease Control. Serum dioxin in Vietnam-era veterans—preliminary report. *MMWR.* 1987;36:470–475.
3. Centers for Disease Control. Serum 2,3,7,8-TCDD levels in Air Force health study participants—preliminary report. *MMWR.* 1988;37:309–311.
4. Gross ML, Lay JO, Lyon PA. 2,3,7,8-Tetrachlorodibenzo-p-dioxin levels in adipose tissue of Vietnam veterans. *Environ Res.* 1984;33:261–268.
5. Kahn PC, Gochfield M, Nygren M, et al. Dioxins and dibenzofurans in blood and adipose tissue of Agent Orange exposed Vietnam veterans and matched controls. *JAMA.* 1988;259:1661–1667.
6. Kang HK, Watanabe KK, Breen JJ, et al. Dioxins and dibenzofurans in adipose tissue of US Vietnam veterans and controls. *Am J Public Health.* 1991;81:344–349.
7. Pirkle JL, Wolfe WH, Patterson DG Jr, et al. Estimates of the half-life of 2,3,7,8-TCDD in Vietnam veterans of Operation Ranch Hand. *J Toxicol Environ Health.* 1989;27:165–171.
8. Schecter A, Constable JD, Bangert JV, et al. Elevated body burdens of 2,3,7,8-tetrachlorodibenzodioxin in adipose tissue of United States Vietnam veterans. *Chemosphere.* 1989;18:431–438.
9. Schecter A, Constable J, Bangert JV, et al. Isomer specific measurement of polychlorinated dibenzo dioxin and dibenzofuran isomers in human blood from American Vietnam veterans two decades after exposure to Agent Orange. *Chemosphere.* 1989;18:531–538.
10. Centers for Disease Control. Preliminary report: 2,3,7,8-TCDD exposure to humans—Seveso, Italy. *MMWR.* 1988;37: 733–736.
11. Facchetti S, Fornari A, Montagna M. Distribution of 2,3,7,8-TCDD in the tissues of a person exposed to the toxic cloud at Seveso. *Adv Mass Spectrom.* 1980;8B:1405–1414.
12. Andrews JS Jr, Garrett WA, Patterson DG Jr, et al. 2,3,7,8-TCDD levels in adipose tissue of persons with no known exposure and in exposed persons. *Chemosphere.* 1989;18:499–506.
13. Needham LL, Allen KM, Patterson DG Jr, Andrews JS Jr. Human levels of 2,3,7,8-tetrachlorodibenzo-p-dioxin in Missouri. *Organohalogen Compounds 1(Toxicol, Environ, Food, Exposure-Risk)* [Bayreuth, Germany: Ecoinforma Press]. 1990;1:259–262.
14. Patterson DG Jr, Needham LL, Pirkle JL, et al. Correlation between serum and adipose tissue levels of 2,3,7,8-tetrachlorodibenzo-p-dioxin in 50 persons from Missouri. *Arch Environ Contam Toxicol.* 1988;17(2):139–143.
15. Patterson DG Jr, Hoffman RE, Needham LL, et al. 2,3,7,8-tetrachlorodibenzo-p-dioxin levels in adipose tissue of exposed and control persons in Missouri—an interim report. *JAMA.* 1986;256:2683–2686.
16. Dewailly E, Tremblay-Rousseau H, Carrier G, et al. PCDDs, PCDFs, and PCBs in human milk of women exposed to a PCB fire and of women from the general population of the province of Québec, Canada. *Organohalogen Compounds 1(Toxicol, Environ, Food, Exposure-Risk)* [Bayreuth, Germany: Ecoinforma Press]. 1990;1:227–230.
17. Ryan JJ, Schecter A, Masuda Y, Kikuchi M. Comparison of PCDDs and PCDFs in the tissue of Yusho patients with those from the general population in Japan and China. *Chemosphere.* 1987;16:2017–2025.
18. Beck H, Eckart K, Mathar W, Wittkowski L. Levels of PCDDs and PCDFs in adipose tissue of occupationally exposed workers. *Chemosphere.* 1989;18:507–516.
19. Patterson DG, Fingerhut MA, Roberts DW, et al. Levels of polychlorinated dibenzo-p-dioxins and dibenzofurans in workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Am J Ind Med.* 1989;16: 135–146.
20. Schecter A, Ryan JJ. Polychlorinated dibenzo-p-dioxin and dibenzofuran levels in human adipose tissues from workers 32 years after occupational exposure to 2,3,7,8-TCDD. *Chemosphere.* 1988;17: 915–920.
21. Graham M, Hileman FD, Orth RG, Wendling JM, Wilson JD. Chlorocarbons in adipose tissue from a Missouri population. *Chemosphere.* 1986;15:1595–1600.
22. Patterson DG Jr, Holler JS, Smith SJ, Liddle JA, Sampson EJ, Needham LL. Human adipose data for 2,3,7,8-TCDD in certain US samples. *Chemosphere.* 1986; 15:2055–2060.
23. Ryan JJ, Schecter A, Lizotte R, Sun WF, Miller L. Tissue distribution of dioxins and furans in humans from the general population. *Chemosphere.* 1985;14:929–932.
24. Ryan JJ, Schecter A, Sun WF, Lizotte R. Distribution of chlorinated dibenzo-p-dioxins and chlorinated dibenzofurans in human tissues from the general population. In: Rappe C, Choudhary G, Keith L, eds. *Chlorinated Dioxins and Dibenzofurans in Perspective.* Chelsea, Mich: Lewis Publishers; 1986.
25. Stanley JS, Bauer KM, Turman K, Boggess K, Cramer P. *Determination of Body Burdens for Polychlorinated Dibenzo-p-dioxins (PCDDs) and Polychlorinated Dibenzofurans (PCDFs) in California Residents.* Springfield, Va: National Technical Information Service; 1989. NTIS publication PB90-148289.
26. *Dioxins and Dibenzofurans in Adipose Tissue of US Vietnam Veterans and Controls.* Washington, DC: Veterans Administration/Environmental Protection Agency; 1990. EPA publication 560/5-89-002. NTIS PB91-16785.
27. Stanley JS, Ayling RE, Cramer PH, et al. Polychlorinated dibenzo-p-dioxins and dibenzofurans in human adipose tissue samples from the continental United States collected from 1972 through 1987. *Chemosphere.* 1990;20:895–901.
28. *NHATS Broad Scan Analysis: Population Estimates from Fiscal Year 1982 Specimens.* Washington, DC: Environmental Protection Agency; 1990. EPA publication EPA 560/5-90-001. NTIS PB90-220567.
29. *Broad Scan Analysis of the FY82 National Human Adipose Tissue Survey Specimens:* vol IV. *Polychlorinated Dibenzo-p-dioxins (PCDD) and Polychlorinated Dibenzofurans (PCDF).* Washington, DC: Environmental Protection Agency; 1986. EPA publication 560/5-86-038. NTIS PB87-177234.
30. Stanley JS, Boggess K, Onstot J, et al. PCDDs and PCDFs in human adipose tissues from the EPA FY82 NHATS repository. *Chemosphere.* 1986;15:1605–1612.
31. *Chlorinated Dioxins and Furans in the General US Population: NHATS FY87 Results.* Washington, DC: Environmental Protection Agency; 1991. EPA publication 560-5-91-003. NTIS PB92-18157.
32. Orban JE, Lordo B. *Statistical Methods for Analyzing NHATS Composite Sample Data—Evaluation of Multiplicative and Additive Methodologies.* Columbus, Ohio: Battelle Memorial Institute draft final report to the Environmental Protection Agency, Office of Pesticides and Toxic Substances, Washington, DC; 1989. Contract 68-02-4243. Document NHATS-SS-11.
33. Kutz FW, Barns DG, Bottimore DP, Greim H, Bretthauer EW. The International Toxicity Equivalency Factor (I-TEF) method of risk assessment for complex mixtures of dioxins and related compounds. *Chemosphere.* 1990;20:751–757.
34. Fingerhut MA, Sweeney MH, Patterson DG, et al. Levels of 2,3,7,8-tetrachlorodibenzo-p-dioxin in the serum of the US chemical workers exposed to dioxin contaminated products: interim results. *Chemosphere.* 1989;19:835–840.
35. Bailar JC. How dangerous is dioxin? *N Engl J Med.* 1991;324:260–262. Editorial.
36. Fingerhut MA, Halperin WE, Marlow DA, et al. Cancer mortality in workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *N Engl J Med.* 1991;324:212–218.

Pesticide Exposures, Cholinesterase Depression, and Symptoms among North Carolina Migrant Farmworkers

Stephen Ciesielski, PhD, MD, Dana P. Loomis, PhD, Susan Rupp Mims, MD, and Annella Auer, MPH

ABSTRACT

Objectives. We conducted a clinic-based study of erythrocyte cholinesterase levels, pesticide exposures, and health effects among farmworkers and nonfarmworkers to determine risks for exposure and associated morbidity.

Methods. Two hundred two farmworkers and 42 nonfarmworkers were recruited sequentially at two community health centers. Erythrocyte cholinesterase levels were measured colorimetrically. Questionnaires obtained data on demographics, occupational history, exposures, and symptoms.

Results. Cholinesterase levels were significantly lower among farmworkers (30.28 U/g hemoglobin) than among nonfarmworkers (32.3 U/g hemoglobin). Twelve percent of farmworkers, but no nonfarmworkers, had very low levels. Farmworkers applying pesticides also had lower cholinesterase levels. One half of farmworkers reported being sprayed by pesticides and working in fields with an obvious chemical smell. Of reported symptoms, only diarrhea was associated with cholinesterase levels. Reported exposures, however, were strongly associated with symptoms.

Conclusions. Farmworkers reported many pesticide exposures that violate state and federal regulations. Farmworkers had cholinesterase levels significantly lower than those of nonfarmworkers, although only spraying pesticides was associated with very low levels. (*Am J Public Health.* 1994;84:446–451)

Introduction

Exposure to cholinesterase-inhibiting pesticides such as organophosphates and carbamates is considered a major health problem for the 2.5 to 5 million farmworkers in the United States and the 80 000 farmworkers in North Carolina.[1–6] Clinical manifestations of such exposure are extremely diverse.[7–9] Long-term exposure has been implicated in several types of cancer, teratogenic effects, sterility, spontaneous abortion, and cognitive deficits.[10,11] Major poisonings among farmworkers are fairly well characterized, but the health effects of routine, smaller exposures are uncertain. Previous studies, although limited by small sample size and inadequate comparison groups, revealed depressed cholinesterase levels in farmworkers,[12–16] the pathophysiological significance of which remains unclear. Thus, we conducted a clinic-based study of farmworkers and nonfarmworkers in North Carolina to assess pesticide exposure, the symptoms occurring among exposed workers, and the effect of simple measures to protect against exposure.

Methods

Subject Selection

Every third patient presenting to two community health centers was an eligible subject. Those performing farmwork for 1 week of the previous month (i.e., those with the potential for occupational exposure) were eligible for the farmworker group. Those who had never performed migrant farmwork were eligible for the unexposed nonfarmworker group. Two hundred two farmworkers and 42 nonfarmworkers participated; nonresponse was 4% (8) for farmworkers and 5% (2) for nonfarmworkers. There was no attrition. Informed consent was obtained from subjects in their preferred language.

Measurement of Symptoms and Pesticide Exposure

Pretested, verbally administered questionnaires obtained information about demographic characteristics, work history, occupational pesticide exposure within 1 month of enrollment, and symptoms within 1 week of enrollment. Nonfarmworkers were questioned about domestic and occupational exposure to pesticides. Because exposure to cholinesterase-inhibiting pesticide produces characteristic biologic effects,[8–10] a fairly comprehensive range of symptoms was included in the questionnaire. Two symptoms thought not to be associated with exposure—increased nail friability and dysgeusia (altered taste perception)—were also included to assess the extent of recall bias.

Erythrocyte Cholinesterase Measurement

Cholinesterase depression represents an unambiguous indication of exposure to cholinesterase-inhibiting pesti-

Stephen Ciesielski is with the Department of Family Medicine at the Valley Medical Center, Fresno, Calif. Dana P. Loomis is with the Department of Epidemiology, and Susan Rupp Mims and Annella Auer are with the Department of Maternal and Child Health, School of Public Health, University of North Carolina, Chapel Hill.

Requests for reprints should be sent to Stephen Ciesielski, PhD, MD, Department of Family Medicine, Valley Medical Center, 445 S Cedar Ave, Fresno, CA 93702.

This paper was accepted August 24, 1993.

TABLE 1—Demographic Characteristics of Study Subjects

	Farm-workers (n = 202)	Non-farm-workers (n = 42)
Male gender, %	77	49
Mean age, y (SE)	33 (0.83)	37 (2.57)
Birthplace, %		
United States	43	52
Mexico	47	36
Central America	6	12
Other	7	0
Mean y in farm-work (SE)	8.2 (0.73)	NA[a]

[a]Not applicable.

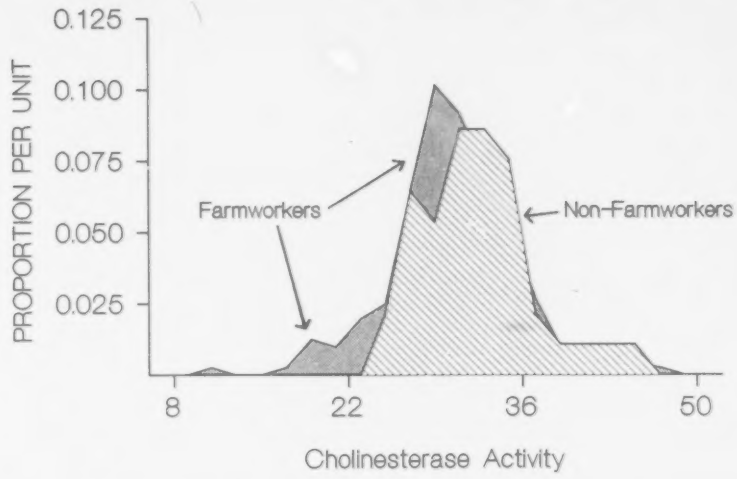


FIGURE 1—Erythrocyte cholinesterase activity of study subjects.

cides as well as a biologic effect of such exposure.

Five microliters of blood were obtained by finger prick. Hemoglobin and acetylcholinesterase activity were measured with the Test-Mate OP Kit (EQM Inc, Junefield, Ohio). This field assay uses a modified Ellman method[17] and has received extensive field and lab testing.[18–21] Acetylcholinesterase activity, measured as absorbance, was corrected for ambient temperature and hemoglobin. Replicate tests were performed on subjects with acetylcholinesterase levels below 25 U/g hemoglobin. Field staff tested themselves daily as an external quality control. Subjects' results from days on which the operator's acetylcholinesterase level varied by more than 10% from the operator's mean were not included in the final analysis.

Symptomaticity may occur only with severe acetylcholinesterase depression.[8] Therefore, measurement of the frequency of adverse effects in a population may depend more strongly on the proportion with very low acetylcholinesterase levels than on those whose levels represent the population mean. To account for this possibility, cholinesterase levels were dichotomized into normal and low categories as well as treated on a continuous scale. The cutpoint for these categories was one standard deviation below the mean of the nonfarmworker group.

Data Analysis

Like many biochemical markers, acetylcholinesterase inhibition can be viewed as a measure of either exposure or outcome. In the first part of the analysis, employment as a migrant farmworker was taken as an indicator of the potential for pesticide exposure, cholinesterase levels were examined among farmworkers and the comparison group of nonfarmworkers, and their mean cholinesterase levels were compared. In addition, odds ratios were used to compare the observed and expected numbers of "cases" with extremely low cholinesterase levels among farmworkers, deriving the expected numbers from the experience of the nonfarmworker comparison group.

Subsequent analyses focused on the relationship of acetylcholinesterase depression among farmworkers to specific occupational and behavioral indicators of potential exposure; nonfarmworkers could not, by definition, have had these exposures. Acetylcholinesterase levels for farmworkers who reported a given exposure or activity were compared with levels for farmworkers without that factor, and the data were treated in both continuous and dichotomous forms as before.

The third and final part of the analysis sought to identify symptoms associated with exposure. Instances of workers who reported a given symptom were treated as cases. Pesticide exposure was defined by both an extremely low acetylcholinesterase level and reported occupational exposure through using pesticides, being sprayed, or working in recently sprayed areas. For each symptom, odds ratios were used to estimate the ratio of observed to expected cases.

The statistical analysis was performed using the SAS system (SAS Institute, Cary, NC). Coefficients for linear regression models were estimated by least squares, with P values derived from standard F tests. Confidence intervals (CIs) for crude odds ratios were based on Woolf's method, and the Mantel-Haenszel chi-squared statistics were used for hypothesis testing.[22] Adjusted odds ratios were estimated by logistic regression.

Results

Both groups were demographically similar with regard to birthplace, but farmworker subjects were younger and more often male (Table 1).

Figure 1 shows the distributions of cholinesterase levels between the two groups. The mean acetylcholinesterase level of the farmworkers (30.18 U/g hemoglobin) was significantly lower than that of the nonfarmworkers (32.20 U/g hemoglobin; $P = .01$) (Table 2). No nonfarmworkers had acetylcholinesterase levels below 25.28 U/g hemoglobin (the lower limit of the nonfarmworker mean minus one standard deviation), while 24 (12%) of the farmworkers did ($P = .019$).

Many farmworkers reported occupational exposure to pesticides. One type of exposure derives from the application of pesticides. Forty-seven percent of

TABLE 2—Erythrocyte Cholinesterase Values among Subjects

	Farmworkers			Nonfarmworkers			
	No.	Mean Value	No. Cases[a]	No.	Mean Value	No. Cases	Difference in Means
Ethnic group							
US-born Blacks	69	31.00	8	11	32.29	0	1.3
Hispanics	120	29.73	13	21	32.17	0	2.4*
Other[b]	13	29.97	3	10	32.18	0	2.2
All subjects	202	30.18	24	42	32.20	0	2.1*

[a]Those with cholinesterase levels below 25.28 U/g hemoglobin, the lower limit of nonfarmworkers (those unexposed) minus one standard deviation.
[b]Haitians, Whites, Puerto Ricans, Jamaicans.
*$P < .05$, Student's t test.

farmworkers reported being sprayed directly with pesticides or working immediately adjacent to rig spraying within 1 month prior to study enrollment (Table 3); 33% of these subjects reported subsequent illness. Fifty-one percent of the subjects reported working in fields with an obvious chemical smell. None of these reported exposures was associated with significantly lower acetylcholinesterase levels although among those reporting exposures was a consistently greater proportion of cases with acetylcholinesterase below 25.28 U/g hemoglobin.

In contrast, the mean acetylcholinesterase level of farmworkers applying pesticides was 2.7 U/g hemoglobin lower than that of other farmworkers ($P = .002$), and the odds ratio for the association of having extremely low acetylcholinesterase with applying pesticides was 4.2 (95% CI = 1.1, 16) (Table 3). There were no significant differences among the mean acetylcholinesterase levels of farmworkers employed in different crops.

A second type of exposure derives from pesticide residues and relates to potentially protective behaviors such as wearing gloves or washing hands (Table 4). Low acetylcholinesterase levels had little or no association with these exposures. The majority of subjects (58%) reported that handwashing water was not consistently available. Similar analyses that used linear and logistic regression models and adjusted for age, years in farmwork, demographic characteristics, and multiple exposures produced essentially the same results.

When the frequency of symptoms among farmworkers was examined with respect to acetylcholinesterase levels below and above 25.28 U/g hemoglobin, only diarrhea was substantially associated with lower levels (odds ratio = 3.1; 95% CI = 1.18, 8.0) (Table 5). However, 10 of the 16 symptoms had positive associations with reported exposure. Neither of the variables included to assess possible recall bias was associated with acetylcholinesterase levels, but dysgeusia did have an association of borderline significance with reported exposure (Table 5).

Discussion

To our knowledge, this is the largest epidemiological study to date of the health effects on farmworkers of exposure to cholinesterase-inhibiting pesticides. A clinic-based design was chosen to maximize inclusion of morbidity in order to examine the biologic relation between illness and acetylcholinesterase levels. The methodological difficulties in the epidemiological study of the health of migrant farmworkers relate primarily to worker transience and workplace conditions. Attrition complicates longitudinal study: in previous research, 74% to 83% of subjects were lost to observation within 1 to 3 years.[23,24] In cross-sectional studies, constraints on determining etiology are compounded by barriers to observational data collection and by the fact that workplace exposure and conditions may change daily.

Reliance on reported exposures poses obvious problems. Most farmworkers do not know to what pesticides they may be exposed. Moreover, although we saw no evidence of it, recall bias may cause overreporting of exposure among symptomatic subjects. Underreporting may be more common; state and federal regulations prohibit fieldwork before sprays have dried or dusts have settled,[25] but perceptible signs of reentry time violation may not exist. An obvious chemical smell indicates that pesticides remain suspended or liquid. Fifty percent of subjects reported exposure both to pesticide suspensions and to aerial or rig spraying, and one third of these subjects reported subsequent illness. Even allowing for substantial misattribution, it appears that occupational exposures to pesticides may routinely result in illness among farmworkers.

The most significant associations between occupational conditions and low acetylcholinesterase levels were demonstrated for workers applying pesticides. North Carolina regulations require that all individuals involved in the application of pesticides receive training and be licensed. No subjects who worked applying pesticides and who had low acetylcholinesterase levels were licensed.

The prevalence of protective behaviors was higher for those actions directly under the workers' control, such as the use of protective clothing. However, even individual protective behaviors such as washing are limited by occupational conditions such as the availability of water. Subjects' motivation to engage in protective behavior is indicated by the fact that as many subjects reported handwashing before eating as reported the availability of handwashing water. The unavailability of handwashing water (reported by 58%) is exactly the same as that obtained by a study conducted several years earlier,[23] suggesting that, despite promulgation of new field sanitation laws, lack of enforcement allows unsafe working conditions to continue.

Associations between exposure and low acetylcholinesterase levels can be examined without knowing the specific pesticides involved because no other significant environmental exposures lower *erythrocyte* cholinesterase.[26] As did earlier studies,[12–16] we found that farmworkers have lower acetylcholinesterase levels than nonfarmworkers. The difference in means was statistically significant, but its biologic meaning is uncertain. Of more probable importance is that 12% of farmworkers had levels below the lower limit of the nonfarmworker group. Since the nonfarmworkers were not really unexposed (all lived in heavily sprayed agricultural areas and many reported using pesticides on gardens or pets), the difference between a farmworker group and an unexposed group may be underestimated.

TABLE 3—Cholinesterase Levels and Occupational Exposures to Pesticide Application Reported by Farmworkers[a]

Exposure Variables Related to Pesticide Application	Farmworkers		Difference in Means of Erythrocyte Cholinesterase Level[b]	Cases[c] among Exposed		Cases among Unexposed		Odds Ratio	95% CI
	No.	%		No.	%	No.	%		
Being sprayed directly or indirectly	92	47	0.1	13	14.1	11	10.5	1.4	0.60, 3.3
Working in fields with chemical odor	99	51	0.1	13	13.1	11	11.6	1.2	0.49, 2.7
Mixing or applying pesticides	9	05	−2.7	3	33.3	1	10.5	4.2	1.1, 16

[a]Reported exposures within the 30 days prior to interview.
[b]Difference between the means of the exposed and unexposed groups.
[c]Subjects with acetylcholinesterase below 25.28 U/g hemoglobin.

The health effects investigated all have multiple causes, many especially prominent in farmwork. Rashes result from plant allergens, dermatoses, sun, heat, infectious causes, and a variety of pesticides. Interactions with other health risks in farmwork may not only complicate diagnosis and study but also place farmworkers at greater risk. For example, dermatoses increase absorption of pesticides, and dehydration and poor nutrition increase their toxic effects.

Only minimal association was observed between low acetylcholinesterase levels and morbidity. However, a moderately strong association was observed with diarrhea, reported as one of the most common effects of such exposure.[27] Several explanations for this relative lack of association exist:

1. No morbidity resulted because acetylcholinesterase levels were not sufficiently depressed to cause appreciable illness. It is also possible that tolerance develops to acetylcholine or that production of acetylcholinesterase increases in the face of repeated exposure.[26]

2. Limitations of cross-sectional study design introduced time-related misclassification of exposure and/or outcomes. Collecting data on exposures in the previous month but on symptoms only during the previous week may have allowed acetylcholinesterase depression to continue but associated symptoms to resolve.

3. Changing exposure to cholinesterase-inhibiting pesticides and other pesticides causing similar symptoms by different mechanisms may obscure the relationship between exposure and illness. Also, since carbamates reversibly inhibit acetylcholinesterase, resolution of previously depressed acetylcholinesterase levels may have occurred before enrollment.

TABLE 4—Cholinesterase Levels and Occupational Exposures to Pesticide Residues Reported by Farmworkers[a]

Exposure Variables Related to Pesticide Residues	Farmworkers		Cases[b] among Exposed		Cases among Unexposed		Odds Ratio	95% CI
	No.	%	No.	%	No.	%		
Aggregate of exposures below	129	66	16	12	8	12	1.0	.42, 2.6
Did not always wear								
Shirt	28	18	3	11	18	14	.73	.20, 2.6*
Long pants	35	18	3	9	21	13	.64	.18, 2.3
Shoes	41	21	5	12	19	12	1.0	.36, 2.9
Long-sleeved shirt	113	57	16	14	8	9	1.6	.67, 4.0
Gloves	165	83	20	12	4	12	1.0	.32, 3.2
Handwashing water not always available[c]	116	58	14	12	10	12	1.0	.43, 2.4
Did not always wash hands								
Before eating in fields	73	50	9	12	8	11	1.1	.41, 3.1
Before smoking in fields[d]	91	96	8	9	0	0	.92	.04, 18.5
Before excretory activity	173	87	20	12	4	14	.76	.24, 2.4
Did not always launder clothing before wearing again	77	38	10	13	14	11	1.2	.49, 2.7
Did not always wash harvested produce before consumption	51	50	4	8	9	18	.38	.12, 1.3

[a]Reported exposures within the 30 days prior to interview.
[b]Subjects with erythrocyte cholinesterase levels below 25.28 U/g hemoglobin.
[c]Availability of handwashing water does not relate directly to residue exposure but directly limits the ability of individuals to accomplish protective behaviors related to the use of water for washing.
[d]For those who smoke.

Reported exposures, however, were strongly associated with reported symptoms. The results of the symptom variables included to assess recall bias suggest that such bias was minimal. Farmworkers who reported exposures were no more likely to report nail friability than those who did not, although dysgeusia was of borderline significance. However, dysgeusia may be indirectly related to exposure since it may result from sialorrhea or nausea.[28]

Given the evidence of exposures in excess of those permitted by federal regulations, of the depression of acetylcholinesterase levels, and of the associa-

TABLE 5—Associations between Reported Symptoms and Low Erythrocyte Cholinesterase Values[a] and Reported Pesticide Exposures[b]

	Low Acetylcholinesterase		Reported Exposures	
	OR[c]	95% CI	OR	95% CI
Symptom/sign				
Diarrhea	3.1	1.18, 8.0	8.2	2.3, 29.2
Nausea	1.8	0.72, 4.2	2.8	1.4, 6.1
Rash	1.5	0.63, 1.9	2.7	1.4, 5.2
Red/irritated eyes	1.5	0.59, 3.6	1.4	0.73, 2.7
Fever	1.6	0.53, 4.2	1.3	0.61, 2.7
Increased sweating	1.4	0.56, 3.4	1.5	0.76, 2.8
Increased anxiety	1.1	0.43, 2.8	2.1	1.5, 6.4
Dizziness	1.2	0.48, 2.9	2.1	1.0, 4.1
Headache	0.91	0.39, 2.1	2.2	1.2, 3.9
Blurred vision	1.4	0.56, 3.4	2.3	1.1, 4.8
Muscular symptoms[d]	0.98	0.41, 2.3	2.8	1.5, 5.2
Chest pain	0.99	0.38, 2.6	4.7	1.7, 13.1
Sialorrhea	0.98	0.31, 3.1	1.3	0.56, 2.8
Difficulty breathing	0.36	0.12, 1.3	1.8	0.87, 3.7
Ataxia	0.55	0.12, 2.4	5.0	1.6, 15.7
Memory loss	0.36	0.05, 2.6	2.9	.87, 10.1
Variables to assess recall bias				
Dysgeusia	1.3	0.43, 3.6	2.5	1.0, 6.0
Friable nails	0.32	0.05, 2.2	1.9	0.67, 5.3

[a]Low erythrocyte cholinesterase defined as below 25.28 U/g hemoglobin.
[b]Category of those exposed includes subjects reporting one of the following: (1) being sprayed directly, (2) working in fields near spraying, (3) working in fields with a chemical smell, or (4) mixing/applying pesticides.
[c]Crude odds ratios were obtained from Mantel-Haenszel chi-squared statistics.
[d]Muscular disorder includes weakness, muscle cramps, and muscle fasciculations.

tions between exposure and reported symptoms, it seems probable that morbidity results from these more routine types of pesticide exposure among farmworkers despite the relative lack of association with acetylcholinesterase depression. It is also possible that the cross-sectional nature of this study limited the identification of morbid sequelae. This conclusion is strengthened by anecdotal experience. The same acetylcholinesterase field assay was later used to diagnose pesticide-related illness in a farmworker clinic. Seventeen individuals with characteristic symptoms were tested; four had acetylcholinesterase levels below the lower limit of normal. Their acetylcholinesterase levels had increased for 3 to 4 successive weeks before the subjects were lost to follow-up; however, their symptoms had usually resolved within 1 week of exposure.

It is important that additional research be undertaken for several reasons. First, migrant health programs have begun to dedicate much of their extremely limited funds to pesticide issues despite uncertainty about the extent of health effects. Most funding goes to educational programs designed to change farmworker behaviors, although exposures leading to serious illness or death are not likely to be under workers' control. Second, this study only addresses acute and subacute neurotoxicity. Thus, the relationship between organophosphate and carbamate exposure and other serious health effects such as cancer remains controversial even though such outcomes may be the most significant health effect of pesticide exposure.[11] Cholinesterase testing may eventually serve a more important role in monitoring workers for long-term cancer risk than for short-term neurotoxicity.

Paradoxically, although pesticides are routinely regarded as a serious health risk for farmworkers, few if any migrant health clinics are capable, in terms of technology, diagnostic protocols, and logistics, of diagnosing pesticide-related illness. We strongly recommend that migrant health centers develop means of routine cholinesterase testing to improve both diagnosis of pesticide poisoning and descriptions of the nature and extent of pesticide-related illness among migrant farmworkers on a population level.

Although population parameters cannot be estimated, some conclusions can nevertheless be drawn from these results. First, many patients presenting to farmworker clinics may have had recent pesticide exposures exceeding those permitted by federal regulations. These patients often believe that such exposures have adversely affected their health. Second, depressed acetylcholinesterase levels are more often found among farmworker patients than among nonfarmworker patients. The clinical significance of these low acetylcholinesterase levels cannot be determined on the basis of these data. Finally, farmworkers' behavior suggests that they are motivated to reduce pesticide exposures but that their ability to do so is constrained by occupational conditions beyond their control. □

Acknowledgments

Funding for this study was provided by the North Carolina Primary Health Care Association (NCPHCA) in Cary, NC. Gratitude is expressed to Susan Sparrow, NCPHCA migrant specialist, to Noni Estelli for assistance with field work, to the Tri-County Community Health Center in Newton Grove, NC, and to Goshen Medical Center in Faison, NC.

References

1. Wilk VA. *The Occupational Health of Migrant and Seasonal Farmworkers in the United States.* Washington, DC: Farmworker Justice Fund; 1986.
2. Moses M. Pesticide-related health problems and farmworkers. *Am Assoc Occup Health Nurs J.* 1989;37:115–130.
3. Lillesand D, Dravitz L, McClellan J. *An Estimate of the Number of Migrant and Seasonal Farmworkers in the United States and the Commonwealth of Puerto Rico.* Washington, DC: Legal Services Corporation; 1977.
4. Fleming MF, Hayak CS. Health care planning in North Carolina: migrant workers program. *N C Med J.* 1984;45:371–374.
5. Coye MJ. The health effects of agricultural production: I. the health of agricultural workers. *J Public Health Policy.* 1985;6:349–370.
6. Kahn E. Pesticide related illness in California farm workers. *J Occup Med.* 1976;18:693–696.
7. Wills JH. The measurement and significance of changes in the cholinesterase activities of erythrocytes and plasma in man and animals. *CRC Crit Rev Toxicol.* 1972;1(March):153–202.
8. Namba T. Cholinesterase inhibition by organophosphorus compounds and its clinical effect. *Bull World Health Organ.* 1971;44:289–307.
9. Midtling JE, Barnett PG, Coye MJ, et al. Clinical management of organophosphate poisoning. *West J Med.* 1985;142:514–518.

10. Rodnitzky RL, Levin HS, Mick DL. Occupational exposure to organophosphate pesticides: a neurobehavioral study. *Arch Environ Health.* 1975;30:98–103.
11. Moses M. Cancer in humans and potential occupational and environmental exposure to pesticides. *Am Assoc Occup Health Nurs J.* 1989;37:131–136.
12. Bodgen JD, Quinones MA, El Nakah A. Pesticide exposure among migrant workers in southern New Jersey. *Bull Environ Contam Toxicol.* 1975;13:513–517.
13. Quinones MA, Bogden JD, Louria DB, El Nakah A. Depressed cholinesterase activities among farmworkers in New Jersey. *Sci Total Environ.* 1976;6:155–159.
14. Richards DM, Kraus JF, Kurtz P, Borhani NO. A controlled field trial of physiological responses to organophosphate residues in farm workers. *J Environ Pathol Toxicol.* 1983;2:493–512.
15. Wicker GW, Williams WA, Guthrie FE. Exposure of field workers to organophosphorus insecticides: sweet corn and peaches. *Arch Environ Contam Toxicol.* 1975;8:175–182.
16. Wicker GW, Williams WA, Bradley JR, Guthrie FE. Exposure of field workers to organophosphorus insecticides: cotton. *Arch Environ Contam Toxicol.* 1975;8:433–440.
17. Ellman GL, Courtney KD, Andres V, Featherstone RM. A new and rapid colorimetric determination of acetylcholinesterase activity. *Clin Chem.* 1987;33:1731–1735.
18. Magnotti RA Jr, Eberly JP, Auarm DEA, McConnell RS. Measurement of acetylcholinesterase in erythrocytes in the field. *Clin Chem.* 1987;33:1731–1735.
19. Magnotti RA, Dowling K, Eberly JP, McConnell RS. Field measurement of plasma and erythrocyte cholinesterases. *Clin Chim Acta.* 1988;315:315–332.
20. McConnell R, Pacheco FA, Magnotti R. Cropduster aviation mechanics: high risk for pesticide poisoning. *Am J Public Health.* 1990;80:1236–1239.
21. Cole D, McConnell R, Murray DL, Pacheco FA. Pesticide illness surveillance; the Nicaraguan experience. *Bull Pan Am Health Organ.* 1988;22:119–130.
22. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research.* New York, NY: Van Nostrand Reinhold; 1982.
23. Ciesielski SD, Seed JR, Ortiz JC, Metts J. Intestinal parasites among North Carolina migrant farmworkers. *Am J Public Health.* 1992;82:1258–1262.
24. Ciesielski SD, Esposito D, Proviva J, Piehl M. The incidence of tuberculosis among North Carolina migrant farmworkers. Unpublished manuscript.
25. Poli B, Fluker S. *Pesticide Safety for Farmworkers.* Washington, DC: Environmental Protection Agency; 1985.
26. Hayes WJ. Pesticides studied in man. Baltimore, Md: Williams & Wilkins Co; 1982.
27. Bardin PG, Van Eeden SF, Joubert Jr: Intensive care management of acute organophosphate poisoning. A 7-year experience in the western Cape. *S Afr Med J.* 1987;72:593–597.
28. Wilson JD, Braunwald E, Isselbacher K, et al., eds. *Harrison's Principles of Internal Medicine.* New York, NY: McGraw Hill; 1991:154.

Public Health Briefs

AIDS Risk Reduction and Reduced HIV Seroconversion among Injection Drug Users in Bangkok

Don C. Des Jarlais, PhD, Kachit Choopanya, MD, MPHTM, Suphak Vanichseni, MD, MPH, Kanokporn Plangsringarm, BSc, Wandee Sonchai, BA, Manuel Carballo, PhD, Patricia Friedmann, MA, and Samuel R. Friedman, PhD

A B S T R A C T

Human immunodeficiency virus (HIV) seroconversion was studied in a group of 173 injection drug users in Bangkok, Thailand, who had been previously tested for HIV and were interviewed and retested in the fall of 1989. Ten percent of the group had seroconverted. Two factors protected against HIV seroconversion: having stopped sharing injection equipment in response to the acquired immunodeficiency syndrome (AIDS) and having a regular sexual partner. The association between self-reported deliberate risk reduction and reduced HIV seroconversion among persons continuing to inject illicit drugs indicates that injection drug users can change their behavior in response to AIDS and will accurately report on the behavior change, and that the changes can protect against HIV infection. (*Am J Public Health* 1994; 84:452–455)

Introduction

Human immunodeficiency virus (HIV) infection has been reported among injection drug users in over 50 different countries.[1,2] Large-scale risk reduction, followed by stabilization of HIV seroprevalence, has been observed among injection drug users in many cities, including New York,[3] Stockholm,[4] San Francisco,[5] Innsbruck,[6] Vienna,[7] and Edinburgh.[8] In all of these cities, however, the association between risk reduction and reduced HIV transmission has been shown only at the community level. To date, there have been no reported studies showing that practicing "safer injection" because of concern about acquired immunodeficiency syndrome (AIDS) reduces the likelihood of HIV seroconversion at the individual level. No study of syringe exchange to date has linked individual use of syringe exchange to a lower risk of HIV seroconversion.[9–13] Similarly, no study has yet shown a protective effect of bleach disinfection against HIV seroconversion at the individual level.[14]

The absence of an individual-level linkage between deliberate risk reduction and reduced HIV seroconversion reinforces the common belief that persons who continue to inject illicit drugs cannot change their AIDS risk behavior and that they will not accurately report their risk behavior. These beliefs have been cited by policymakers in the United States as a rationale for opposing AIDS prevention programs that involve legal access to sterile injection equipment.[15,16]

We report here evidence that deliberate AIDS risk reduction behavior has a protective effect on HIV seroconversion at the individual level among persons continuing to inject illicit drugs in Bangkok, Thailand. Transmission of HIV among injection drug users in Bangkok had been among the most rapid in any population. In studies of injection drug users attending drug abuse treatment clinics in Bangkok, seroprevalence increased from 1% in late 1987 to 15% in March 1988 to 43% in the fall of 1988.[17,18] As public health officials became aware of the spread of HIV among injection drug users in Bangkok, AIDS education/prevention programs were rapidly implemented, including prominent posters in drug abuse treatment clinics. Staff at these clinics encouraged injection drug users not to "share" injection equipment and to purchase sterile injection equipment at pharmacies (prescriptions are not required); educated injection drug users about

Don C. Des Jarlais and Patricia Friedmann are with the Beth Israel Medical Center, New York, NY. Kachit Choopanya, Suphak Vanichseni, Kanokporn Plangsringarm, and Wandee Sonchai are with the Office of the Permanent Secretary, Bangkok Metropolitan Administration (BMA), Bangkok, Thailand. Manuel Carballo is with the Program on Substance Abuse, World Health Organization, Geneva, Switzerland. Samuel R. Friedman is with National Development and Research Institutes, Inc, New York, NY.

Requests for reprints should be sent to Don C. Des Jarlais, PhD, c/o BIMC/NDRI, 11 Beach St, New York, NY 10013.

This paper was accepted October 25, 1993.

Note. The views expressed in this article do not necessarily reflect the positions of the granting agencies or of the institutions by which the authors are employed.

Editor's Note: See related annotation by Chitwood (p 350) in this issue.

TABLE 1—Selected Demographic and Drug Use Characteristics of Bangkok Injection Drug Users (n = 173) Who Previously Tested HIV Negative

	No.	%
Gender		
Male	163	94
Female	10	6
Ethnicity		
Thai	172	99
Other	1	1
Jail history since time of first injection[a]		
Yes	121	71
No	50	29
Frequency of giving, lending, or selling injection equipment in previous 6 months[a]		
Never	77	45
Less than once per week	66	39
One or more times per week	28	16
Stopped sharing injection equipment since hearing about AIDS		
Yes	111	64
No	62	36
Injected heroin	173	100
Injected other drugs in combination with heroin	11	6

Note. The average age of the subjects was 31.54 years (SD = 8.56 years).
[a]Data are missing for 2 subjects.

TABLE 2—Factors Associated with HIV Seroconversion among Injection Drug Users in Bangkok

	Seroconverted (n = 17)		Did Not Seroconvert (n = 156)		
	No.	%	No.	%	$P(\chi^2)$
Incarceration since time of first injection					
Yes	16	94	105	68	
No	1	6	49	43	.026
Stopped sharing injection equipment since hearing about AIDS					
Yes	6	35	105	67	
No	11	65	51	33	.009
Number of primary partners in the past 6 months					
None	14	82	85	54	
One or more	3	18	71	46	.027

sexual HIV transmission; distributed condoms; and provided HIV testing and counseling. Since the fall of 1988, seroprevalence among in-treatment injecting drug users has stabilized: 45% were HIV positive in the spring of 1989[18] and 39% were HIV positive in the fall of 1989.[19] In the fall 1989 study, 92% of the subjects reported that they had changed their behavior to reduce the risk of developing AIDS, with 59% reporting that they had "stopped sharing" injecting equipment because of concern about AIDS.

Methods

In the fall of 1989, 601 injection drug users were recruited at the clinics of the Bangkok Metropolitan Administration drug abuse treatment system. Injecting within the past 2 months was required for study eligibility. Subjects were selected randomly from among persons who had been in treatment for at least 3 months (n = 342) and randomly from among persons who were entering the treatment system (n = 259). All persons asked to participate agreed to do so. At that time, the primary treatment regimen provided in these clinics was a 45-day methadone-assisted detoxification program, followed by aftercare counseling. (Methadone maintenance programming in Bangkok has since been greatly expanded.)

Informed consent was obtained, a risk behavior questionnaire was administered by trained interviewers, HIV test counseling was conducted, and a blood sample was obtained. The questionnaire was developed as part of a multisite World Health Organization study and emphasizes risk behavior in the 6-month period before the interview. Sera were tested for anti-HIV antibodies by double ELISA tests (Wellcome) with confirmation of positive results by Western blot assay (Pacific Health Care) at the Siriraj Hospital laboratory.

Two hundred sixty-three subjects reported that they had previously been tested for HIV, with 173 reporting that their most recent previous test was negative. Retesting of these 173 presented an opportunity to identify risk factors for HIV seroconversion.

Chi-square tests and multiple logistic regression (SAS statistical programs[20]) were used to test associations with HIV seroconversion.

Results

Table 1 presents selected demographic and drug history characteristics of the 173 subjects reporting a previous HIV-negative test. The mean time between most recent previous test and current test was 9.8 months (SD = 7.2; range = 1–32, median = 8).

Seventeen (10%) subjects had seroconverted by the fall of 1989. Bangkok Metropolitan Administration records were used to validate the previous HIV-negative test for all of these 17 subjects. Assuming that the seroconversions occurred midway between the most recent negative HIV test and the fall 1989 test gave an estimated HIV seroconversion rate of 12.5 per 100 person-years at risk for this sample (95% confidence interval [CI] = 7.6, 17.4).

The 17 subjects who had seroconverted were compared with the 156 subjects who remained seronegative on demographic, drug use, and sexual behavior variables. As shown in Table 2, the three variables that significantly distinguished the seroconverters were (1) having been incarcerated since starting drug injection (a risk factor); (2) reporting that they had "stopped sharing injection equipment in response to AIDS" (a protective factor); and (3) having a primary sexual partner (a protective factor) (a "primary" partner was defined for the subjects as a "regular, most important" sexual partner). The "stopped sharing" variable was constructed from an open format question: "Since you first heard about AIDS, have *you* done anything to avoid catching the virus yourself, or to prevent

TABLE 3—Multivariate Logistic Regression Analysis for HIV Seroconversion among Injection Drug Users in Bangkok

Factor	χ^2	*P*	Odds Ratio (95% CI)
Did not stop sharing injection equipment	6.475	.01	4.01 (1.38,11.66)
Lacks primary sexual partner	6.589	.03	4.17 (1.13,15.41)

Note. CI = confidence interval.

someone else getting it from you?" Subjects were coded as "having stopped sharing" if they responded to this question by reporting that they had "stopped sharing" injection equipment and if they also responded on the follow-up question that they had "maintained this behavior change."

Multiple logistic regression with backwards elimination was used to examine the statistical independence of the three factors associated with HIV seroconversion. (The wording for the two protective factors was reversed so that they were entered as risk factors.) Results are presented in Table 3; failure to stop sharing and lacking a primary sexual partner remained statistically significant.

Most of the risk behavior questions in the questionnaire asked about behavior in the previous 6 months, although many of the seroconversions may have occurred earlier. We therefore repeated the analyses using only the 103 subjects (10 seroconverters) who had tested negative within the 10 months before the fall 1989 testing (and thus presumably had seroconverted within the previous 5 months). To assess the possibility that recently tested nonconverters did not have sufficient time for seroconversion, the analyses were also repeated omitting the 19 subjects (1 seroconverter) who reported that their previous HIV-negative test was within 3 months before the fall 1989 HIV testing. The same pattern of relationships was found in these two subanalyses as for the 173 subjects. (Data available from the corresponding author.)

Discussion

In this study, injection drug users who reported practicing safer injection ("stopped sharing injection equipment") because of concern about AIDS were less likely to have seroconverted for HIV. The time period (1987 through 1989) covered by this study was a period of high rates of both new HIV infections and AIDS risk reduction among injection drug users in Bangkok. Concurrence of high rates of seroconversion and deliberate risk reduction may be an important factor in our ability to show a statistical association between the two processes.

Although lack of documentation of previous HIV tests for the nonconverters is a limitation of this study, the high accuracy of the test reports of the seroconverters and the generally high accuracy of reported HIV test results in another study of injection drug users[21] suggest that few subjects would report a previous test without having been tested.

The 12.5 seroconversions per 100 person-years at risk should *not* be considered as representative of the seroconversion rate among all injection drug users in Bangkok from 1987 through 1989. All subjects in this study would have known about AIDS from their previous HIV counseling and testing and thus would have had some opportunity to change their behavior before HIV seroconversion (although not necessarily before HIV infection). Given the rapidity of the spread of HIV among injection drug users in Bangkok,[19] it is likely that most of the individuals who became infected between 1987 and 1989 did so before they learned about AIDS.

It is not surprising that no specific sexual behavior was associated with HIV seroconversion. The great majority of injection drug users in Bangkok are male, and few have sexual partners who are injection drug users. At the time of this data collection, there were few HIV-infected persons other than injection drug users in the city.

Having a regular sexual partner—an indicator of social integration—had a protective effect against seroconversion in this study. A number of previous studies of injection drug users found that lack of social integration was associated with higher rates of HIV exposure and/or higher rates of risk behavior: being part of an ethnic minority group; being homeless; having a particularly low income; and residing in a shelter or boardinghouse.[22-26] The mechanisms through which these indicators of social marginalization are related to higher rates of risk behavior or higher rates of HIV infection have not been identified, but such relationships do exist in different national settings.

Reporting that one had stopped sharing drug injection equipment in response to AIDS was strongly associated with not seroconverting in this study. To our knowledge, this is the first time that self-reported deliberate AIDS risk reduction among persons continuing to inject drugs has been associated at the individual level with a reduced likelihood of new infection with HIV. These data provide evidence for the construct validity of self-reported AIDS risk reduction among injecting drug users. (The "validity" of a technique for measuring behavior or psychological/social characteristics refers to whether the technique actually measures what it purports to measure.[27] Construct or "predictive" validity is demonstrated when the measured behavior/characteristic is statistically associated with other behaviors/characteristics according to an explicit theoretical framework. Construct validity for measurement of AIDS risk behavior of injection drug users has been demonstrated through the many studies showing strong associations between self-reported risk behavior and HIV status. Valid measurement of behavior change is, however, a much more difficult task than valid measurement of recent behavior, so that it is important to separately demonstrate validity for self-report measurement of AIDS-related behavior changes.) This finding serves to increase confidence that injection drug users can and will change their behavior to reduce the chances of developing AIDS, that they can and will accurately report such changes in behavior, and that such changes in behavior can directly protect the individual drug injector against infection with HIV. □

Acknowledgments

This research was supported by grant DA03574 from the National Institute on Drug Abuse, as well as by the Bangkok Metropolitan Administration and the World Health Organization.

We thank Thomas Ward for editorial assistance and expertise during the preparation of this manuscript.

References

1. Stimson GV. The global diffusion of injecting drug use: implications for HIV infection. Presented at the Eighth International Conference on AIDS; July 1992; Amsterdam, the Netherlands.
2. Des Jarlais DC, Friedman SR, Choopanya K, Vanichseni S, Ward TP. International epidemiology of HIV and AIDS among injecting drug users. *AIDS*. 1992;6: 1053–1068.
3. Des Jarlais DC, Friedman SR, Novick D, et al. HIV-1 infection among intravenous drug users in Manhattan, New York City. *JAMA*. 1989;261:1008–1012.
4. Kall K, Olin R. HIV status and changes in risk behavior among intravenous drug users in Stockholm 1987–88. *AIDS*. 1990; 4:153–157.
5. Watters JK, Cheng Y, Segal M, et al. Epidemiology and prevention of HIV in intravenous drug users in San Francisco, 1986–1989. Presented at the Sixth International Conference on AIDS. June 22, 1990; San Francisco, Calif.
6. Fuchs D, Unterweger B, Hausen A, et al. Anti-HIV-1 antibodies, anti-HTLV-1 antibodies and neopterin levels in parenteral drug addicts in the Austrian Tyrol. *J AIDS*. 1988;1:65–66.
7. Loimer N, Presslich O, Hollerer E, Pakesch G, Pfersman V, Werner E. Monitoring HIV-1 infection prevalence among intravenous drug users in Vienna 1986–1990. *AIDS Care*. 1990;2:281–286.
8. Robertson JR, Skidmore CA, Roberts JJK. HIV infection in intravenous drug users: a follow-up study indicating changes in risk-taking behaviour. *Br J Addict*. 1988;83:387–391.
9. US General Accounting Office. *Needle Exchange Programs: Research Suggests Promise as an AIDS Prevention Strategy*. Report to the Chairman, Select Committee on Narcotics Abuse and Control. Washington, DC: United States House of Representatives; 1993.
10. Lart R, Stimson GV. National survey of syringe exchange schemes in England. *Br J Addict*. 1990;85:1433–1443.
11. Stimson GV. Risk reduction by injecting drug users with regard to HIV infection. *Int Rev Psychiatry*. 1991;3:401–415.
12. Des Jarlais DC, Friedman SR. The AIDS epidemic and legal access to sterile equipment for injecting illicit drugs. *Ann Am Acad Polit Soc Sci*. 1992;521:42–65.
13. Des Jarlais DC, Friedman SR. Research. In: J. Stryker, ed. *Dimensions of HIV Prevention: Needle Exchange*. Menlo Park, Calif: Henry J. Kaiser Family Foundation; 1993:63–75.
14. Curran JW, Scheckel LW, Millstein RA. *HIV/AIDS Prev Bull*. (Atlanta, Ga: National Institute on Drug Abuse, Center for Substance Abuse Treatment.) April 19, 1993.
15. Pear R. U.S. drug official urges mayors to forgo needle-swap programs. *The New York Times*. June 3, 1992:B3.
16. Anderson W. The New York needle trial: the politics of public health in the age of AIDS. *Am J Public Health*. 1991;81:1506–1517.
17. Phanuphak P, Poshyachinda V, Unkeklabh T, Rojanapithayakorn W. HIV transmission among intravenous drug abusers. Presented at the Fifth International Conference on AIDS; Quebec, June 1989; Montreal, Quebec, Canada. Abstract T.G.O.25.
18. Choopanya K, Vanichseni S, Plangsringarm K, et al. Risk factors and HIV seropositivity among injecting drug users in Bangkok. *AIDS*. 1991;5:1509–1513.
19. Vanichseni S, Sakuntanaga P. Results of three seroprevalence studies for HIV in IVDU in Bangkok. Presented at the Sixth International Conference on AIDS; June 1990; San Francisco, Calif. Abstract F.C.105.
20. SAS Institute Inc. *SAS User's Guide: Statistics, Version 5 Edition*. Cary, NC: SAS Institute Inc; 1985.
21. McCusker J, Stoddard AM, McCarthy E. The validity of self-reported HIV antibody test results. *Am J Public Health*. 1992;82:567–569.
22. Friedman SR, Jose B, Deren S, et al. Preliminary analysis of HIV seroconversions among drug injectors in 13 cities. *Proceedings of the Third Annual National AIDS Demonstration Research Conference, Bethesda, Md, 1991*. Rockville, Md: National Institute on Drug Abuse. In press.
23. Kouzi AC, Des Jarlais DC, Tross S, et al. Contraceptive behavior among intravenous drug users at risk for AIDS. *Pyschol Addict Behav*. 1992;6:135–139.
24. Schoenbaum EE, Hartel D, Selwyn PA, et al. Risk factors for human immunodeficiency virus infection in intravenous drug users. *N Engl J Med*. 1989;321:874–879.
25. Selik RM, Castro KG, Pappaioanou M. Racial/ethnic differences in the risk of AIDS in the United States. *Am J Public Health*. 1988;78:1539–1545.
26. van den Hoek JAR, Coutinho RA, van Haastrecht HJA, et al. Prevalence and risk factors of HIV infections among drug users and drug-using prostitutes in Amsterdam. *AIDS*. 1988;2:55–60.
27. Thorndike RL. *Applied Psychometrics*. Boston, Mass: Houghton Mifflin; 1982: 186–192.

Epidemiology Summer Program to Be Held at Tufts

The New England Epidemiology Institute will sponsor the 14th annual Epidemiology Summer Program, July 11 through July 29, 1994, at Tufts University's Medford, Mass, campus. The program includes methodological, statistical, and substantive courses intended for those seeking an introduction to modern epidemiological concepts, as well as for those desiring a review of recent developments in epidemiological thinking.

This year the Institute will be offering 1- and 2-week sessions, allowing participants to attend for 1, 2, or 3 weeks. The 2-week session will run from July 11 through July 22 and will include the theory and practice of epidemiology (introductory and advanced levels), basic biostatistics, regression methods, survival analysis, genetic epidemiology, nutritional epidemiology, epidemiology in developing countries, and clinical research. The 1-week session will run from July 25 through July 29 and will include an introduction to epidemiology, conducting epidemiological research, pharmacoepidemiology, epidemiological research in women's health, the biology and epidemiology of cancer, epidemiology in public health practice, perinatal epidemiology, and occupational epidemiology.

Invited faculty include teachers and researchers from leading universities. Registrants may receive graduate-degree credit or continuing education credits from Tufts University, Continuing Medical Education (AMA Category 1) through Tufts University Medical School, or Certification Maintenance from the American Industrial Hygiene Association.

For more information please contact the New England Epidemiology Institute, Dept PA-AJPH, One Newton Executive Park, Newton Lower Falls, MA 02162-1450; tel (617) 244-1200; fax (617) 244-9669.

Racial Differences in Rural Adults' Attitudes toward Issues of Adolescent Sexuality

Ronnie D. Horner, PhD, Kathryn M. Kolasa, PhD, RD, Thomas G. Irons, MD, and Kenneth Wilson, PhD

ABSTRACT

This study, based on a random sample of adults in a rural North Carolina county, demonstrates racial differences in rural adults' attitudes relating to adolescent sexual issues. Blacks were 50% more likely than Whites to indicate that public schools should provide general health care services, including pregnancy testing and treatment of sexually transmitted diseases, to teenagers; however, they were only half as likely as Whites to approve of sexual experimentation by adolescents. The local community's attitudes must be considered in the implementation of rural adolescent health programs, including acquired immunodeficiency syndrome education. (*Am J Public Health.* 1994;84:456–459)

Introduction

An increasing emphasis is being given to local community initiatives, such as public school–based programs, for educating adolescents about and assisting them to avoid the risks associated with premarital sexual activity, particularly pregnancy and infection with sexually transmitted diseases such as the human immunodeficiency virus (HIV).[1,2]

Adults' attitudes toward adolescent sexual activity and intervention programs influence the content and success of these programs.[3–7] However, this knowledge derives almost exclusively from urban studies. The few studies of rural populations have focused on the extent of adolescent sexual activity rather than adults' attitudes regarding such activity and possible intervention programs.[8–10] This report presents the results of a survey on attitudes regarding adolescent sexuality and related issues held by Black and White adults in a rural community.

Methods

This is a secondary analysis of data that were originally collected for an investigation of rural, southern adults' attitudes and beliefs regarding adolescent health issues, including adolescent sexual activity and pregnancy, sex education, and the provision of health services to teenagers through the county public schools. The survey was the first phase of a community-based intervention program aimed at reducing the number of teenage pregnancies in this rural North Carolina county, which has one of the highest rates of adolescent pregnancy in the state.[11]

The respondents were a random sample of adults (aged 18 years and older) stratified on age and sex. Telephone or in-person interviews were completed for 835 of the 1037 sampled households (80.5%). There were no systematic differences in responses based on type of interview, and respondents were sociodemographically representative of the county population. The interview schedule was pretested and revised until it was effective with respondents from all social backgrounds.[12]

The data analysis was based on the 801 respondents who were either White or Black. Individual questions and indexes composed of related questions (see Appendix) were analyzed. The data were summarized by proportions and means with standard deviations, and racial differences in responses were analyzed by chi-square tests or *t* tests. When significant racial differences were detected, the effect of race was assessed by logistic regression or analysis of covariance (for the indexes) adjusting for covariates shown by preliminary analyses to be potential modifiers of self-reported attitudes, including sex, age, marital status (married vs other), income, education (college vs other), place of residence (central community vs surrounding communities), political orientation (conservative vs other), and religiosity (attend church at least once a week vs less frequently). Because the unadjusted and adjusted results were consistent with few exceptions, which are noted, the former are presented for simplicity.

Results

Characteristics of Respondents

Black and White respondents differed in their sociodemographic charac-

Ronnie D. Horner is with the Department of Medicine, Duke University Medical Center and the Center for Health Services Research in Primary Care, Veterans Affairs Medical Center, Durham, NC. Kathryn M. Kolasa is with the Department of Family Medicine, and Thomas G. Irons is with the Department of Pediatrics and Office of the Dean, East Carolina University School of Medicine, Greenville, NC. Kenneth Wilson is with the Department of Sociology, East Carolina University.

Requests for reprints should be sent to Kathryn M. Kolasa, PhD, RD, Department of Family Medicine, School of Medicine, East Carolina University, Greenville, NC 27858.

This paper was accepted June 23, 1993.

teristics in the expected directions. Blacks were more likely to be women, to be unmarried, to have lower income, and to attend church at least weekly, but they were less likely to have attended college or to live in the central community of the county.

Provision of Health Services to Adolescents

Blacks and Whites agreed equally that public schools should provide health services to adolescents but differed significantly on the types of services to be provided (Table 1). Black respondents were about 50% more likely than Whites to believe that general health services and health services relating to sexual activity (e.g., pregnancy testing) should be provided. The indexes also indicated stronger support for these services by Blacks. The effect of race disappeared with adjustment for sociodemographic characteristics, political orientation, and religiosity only for provision of birth control products to girls, provision of child care instruction, and the Assistance with Personal Problems Index. There were no significant racial differences regarding provision of educationally related health services, such as the distribution of birth control information.

Sex Education for Adolescents

No race effect existed regarding the delivery of sex and acquired immunodeficiency syndrome (AIDS) education to teenagers through the public schools (Table 2). Almost all respondents believed that sex and AIDS education should begin before high school, with about 50% indicating that instruction should begin in middle school (i.e., grades 6 to 8). Only 60% of either race believed that sex education was likely to decrease the number of teenage pregnancies.

Adolescent Sexual Activity

Black respondents held significantly more conservative beliefs toward teenage sexual experimentation than did White respondents (Table 3). Approximately 20% of Black adults agreed that it was acceptable for male or female teenagers to engage in premarital sexual intercourse, in comparison with more than 40% of White adults. Two thirds of Blacks and almost 60% of Whites believed that they, personally, should have been better prepared about sex when they were teenagers.

TABLE 1—Community Attitudes toward the Provision of Health Services in Public Schools and Types of Services to Be Provided

	Black (n = 200), %	White (n = 601), %	*P*
Provision of health services			
Health services should be provided at junior high (middle) school	97.0	94.8	.21
Health serices should be provided at high school	98.0	94.8	.09
Types of services to be provided			
Preventive Medicine Index	14.4	18.5	<.001
Major injury treatment	66.3	40.7	<.001
Physical examinations	70.4	42.5	<.001
Vaccinations	77.6	58.6	<.001
Female checkups	59.2	29.1	<.001
Treatment for acne and other usual problems of adolescents	59.2	35.6	<.001
Pregnancy testing	61.7	42.8	<.001
Diagnosis for sexually transmitted diseases	65.3	48.2	<.001
Tests for chronic diseases	80.1	60.0	<.001
Blood tests	76.9	60.0	<.001
Assistance with Personal Problems Index	7.1	7.7	<.001
Information on good diets	95.5	96.4	.78
Weight loss program	84.3	80.4	.27
Mental health counseling	94.9	91.7	.09
Instruction on how to care for children	90.9	83.9	.06
Counseling for eating disorders	85.8	79.3	.13
"Quit smoking" classes	92.4	81.7	<.001
Assistance with Sexual Problems Index	6.4	7.3	<.001
Birth control products for girls	54.1	42.8	.01
Birth control products for boys	53.1	47.7	.23
Counseling for pregnant adolescents	87.2	88.8	.66
Day care for infants of teenage mothers	79.6	53.1	<.001
Preventive Education Index	3.4	3.3	.28
Birth control information to girls	91.8	93.3	.59
Birth control information to boys	89.2	92.5	.22
Drug and alcohol education	95.4	98.3	.02
Instruction on child care	90.8	84.0	.03

Note. For indexes, means rather than percentages are presented.

Black respondents were also more conservative regarding abortion, being about half as likely as Whites to agree that abortion should be allowed (Table 4). Blacks were less likely to believe that any specific circumstance, such as age of the mother, warranted abortion; the higher Support for Abortion Index score for Blacks reflects this more negative view. After adjustment for sociodemographic characteristics, political orientation, and religiosity, Blacks remained significantly less favorable toward abortion in general and when the pregnancy was due to rape, although no race effect was found for the Support for Abortion Index. Blacks were three times as likely (13.5% vs 4.3%) and about six times as likely (25.5% vs 4.3%) as Whites to believe that an unmarried teenager should either marry the father or keep the baby, respectively.

Discussion

This study is the first description of racial variations in *rural* adults' attitudes regarding adolescent health and sexuality issues. Our findings suggest that Blacks in this rural community are more likely than Whites to support public school–based, comprehensive adolescent health programs. While there was strong support for sex education programs in general, Black adults were more accepting of programs that would also provide general health care and care related to sexuality (e.g., pregnancy testing). We conclude that Black adults of this county are more supportive of adolescent health programs offering a broad spectrum of options, ranging from avoidance of pregnancy through abstinence to school-based day care for infants of teenage mothers. Abortion, however, may be a

TABLE 2—Knowledge and Attitudes of Respondents Regarding AIDS and Sex Education in Public Schools

Area of Knowledge or Attitude	Black (n = 200), %	White (n = 601), %	P
Sex education			
Public schools should provide sex education	91.5	90.7	.84
Grade level for teaching sex education			.94
Elementary school	38.0	37.1	
Middle school	54.2	54.2	
High school or later	7.8	8.6	
Effect of sex education on adolescent pregnancy rate			.37
Increase pregnancies	8.5	7.0	
Decrease pregnancies	57.5	60.7	
No effect	23.0	24.8	
Unsure of effect	11.0	7.5	
AIDS education			
Schools should teach AIDS prevention	98.0	96.7	.48
Grade level for teaching about AIDS			.44
Elementary school	39.3	35.2	
Middle school	51.3	52.7	
High school or later	9.4	12.1	
Formally educated on AIDS	34.5	38.4	.36
Level of AIDS knowledge			<.001
A lot	19.5	36.8	
Some	30.5	40.8	
Little	30.0	16.1	
Nothing	20.0	6.3	

TABLE 3—Attitudes of Respondents toward Their Own and Others' Adolescent Sexuality

Attitude toward Sexuality	Black (n = 200), %	White (n = 601), %	P
Should have been better prepared about sex as a teenager	66.5	58.4	052
Okay for boys to experiment with premarital sex	23.0	45.4	<.001
Okay for girls to experiment with premarital sex	20.5	42.4	<.001

TABLE 4—Attitudes of Respondents Regarding Circumstances for Abortion

Circumstance for Abortion (Support for Abortion Index[a])	Black (n = 200), %	White (n = 601), %	P
Abortions should be allowed in general	46.5	79.5	<.001
Circumstance that warrants abortion			
If pregnancy is the result of rape	69.0	86.7	<.001
If pregnancy will be economic hardship	33.0	44.3	.015
If pregnancy is harmful to woman's health	79.0	89.9	<.001
If pregnancy is unwanted	31.0	40.8	.026
If pregnancy is due to not using birth control	17.0	28.8	.004
If pregnancy is due to birth control failure	27.5	38.1	.025
If the woman is unmarried	21.0	34.9	<.001
If the woman is a teenager	21.5	38.9	<.001
If the woman is over 40	27.0	43.8	<.001

[a]Black, mean = 19.4; White, mean = 17.1 ($P < .001$).

less acceptable option for adults residing in this rural Black community. These findings highlight the need to consider communities individually when designing adolescent health programs. □

Acknowledgments

This research is part of the Pitt County Adolescent Health Project, which is funded in part by the Kellogg Foundation.

We gratefully acknowledge the Survey Research Center of the East Carolina University Regional Development Institute for its assistance in sampling and data collection.

References

1. Centers for Disease Control. Sexual behavior among high school students—United States, 1990. *MMWR.* 1992;40:885–888.
2. Centers for Disease Control. Selected behaviors that increase risk for HIV infection among high school students—United States, 1990. *MMWR.* 1992;41:231, 237–240.
3. Durant RH, Jay S, Seymore C. Contraceptive and sexual behavior of Black female adolescents. A test of a social-psychological theoretical model. *J Adolesc Health Care.* 1990;11:326–334.
4. Thorton A, Camburn D. The influence of the family on premarital sexual attitudes and behavior. *Demography.* 1987;24:323–340.
5. Furstenberg FF. Race differences in teenage sexuality, pregnancy, and adolescent childbearing. *Milbank Q.* 1987; 65(suppl):381–403.
6. Furstenberg FF, Moore KA, Peterson JL. Sex education and sexual experience among adolescents. *Am J Public Health.* 1985;75:1331–1332.
7. Strunin L, Hingson R. Acquired immunodeficiency syndrome and adolescents: knowledge, beliefs, attitudes, and behaviors. *Pediatrics.* 1987;79:825–828.
8. Alexander CS, Ensminger ME, Kim YJ, Smith BJ, Johnson KE. Early sexual activity among adolescents in small towns and rural areas: race and gender patterns. *Fam Plann Perspect.* 1989;21:261–266.
9. Yawn BP, Yawn RA. Teen-age sexual activity in rural Minnesota. *Minn Med.* 1987;70:38–39.
10. McCormick N, Izzo A, Folcik J. Adolescents' values, sexuality, and contraception in a rural New York county. *Adolescence.* 1985;20:385–395.
11. *North Carolina Vital Statistics, 1989.* Raleigh, NC: North Carolina Department of Health and Environmental Resources; 1990.
12. Wilson KR. *Support for Expanded High School Health Services in a Rural Southern County: Assessing the Effects of Regional Subculture and Social Stratification.* Cincinnati, Ohio: American Sociological Association; 1992.

APPENDIX—Indexes of Related Questions

Factor analysis was used to identify sets of related questions from which indexes were created. The 24 questions on health services (Table 1) fell into five indexes, of which four had acceptable levels of reliability.

The Preventive Medicine Index consisted of 9 items, including questions about provision of female checkups, physical examinations, and blood tests (Cronbach alpha = .86; mean = 17.4, SD = 5.9).

The Assistance with Personal Problems Index was composed of 6 items focusing on health services such as counseling for eating disorders, instruction on the care of children, and mental health counseling (Cronbach alpha = .72, mean = 7. , SD = 2.5).

The Assistance with Sexual Problems Index was composed of 4 items including distribution of birth control products to girls and boys, counseling for pregnant teenagers, and day care for infants of teenagers (Cronbach alpha = .73; mean = 7.0, SD = 2.7).

The Preventive Education Index consisted of 3 items, including questions about distributing birth control information to girls and boys and provision of drug and alcohol education (Cronbach alpha = .76, mean = 3.3, SD = 1.1).

Because the items presented in Tables 2 and 3 yielded five factors of two questions each, the indexes were no more informative than the individual items and, hence, were not analyzed. The questions on abortion (Table 4) were combined into the Support for Abortion Index (Cronbach alpha = .92, mean = 17.4, SD = 6.1).

The Weights and Heights of Mexican-American Adolescents: The Accuracy of Self-Reports

Harold Davis, MD, and Peter J. Gergen, MD, MPH

ABSTRACT

The accuracy of Mexican-American adolescents' self-reported weights, heights, and body mass indexes was evaluated with data from the Hispanic Health and Nutrition Examination Survey. On average, adolescents with low measured body mass indexes and high measured body mass indexes overestimated and underestimated their weights, respectively. Categories of low and high body mass indexes created by applying cutoffs to reported body mass indexes had low sensitivities. For weight, height, and body mass indexes, measured and reported values were highly correlated. This high correlation suggests that adolescents' reported values can be used as continuous variables in multivariate analyses with only small errors resulting in the coefficients for weight, height, and body mass index. (*Am J Public Health*. 1994;84:459–462)

Introduction

Researchers wanting to collect data on survey participants' weights and heights might find it desirable to obtain such data by self-report. Studies of reported weights and heights usually have focused on adults[1–13]; fewer studies[14–19] have focused on adolescents. To evaluate the accuracy of Mexican-American adolescents' self-reported weights and heights, we used data from the Hispanic Health and Nutrition Examination Survey.

Materials and Methods

Study Design of the Hispanic Health and Nutrition Examination Survey

The Hispanic Health and Nutrition Examination Survey was conducted by the National Center for Health Statistics from 1982 through 1984. The Mexican-American portion was done from July 1982 through December 1983 among civilian, noninstitutionalized persons of Mexican origin or ancestry in Arizona, California, Colorado, New Mexico, and Texas. The sample design was a stratified, multistage probability sample.[20] The weighted estimates presented are representative of Mexican Americans in these states.

The survey included an interview and an examination. For persons who were 12 through 17 years old, either self-response or response by proxy was acceptable. Persons who were 18 years old or older responded for themselves unless they were unable to be interviewed. In the interview, weights (without shoes) were reported in pounds and heights (without shoes) in feet and inches. In the examination (done a few weeks after the interview), weights to the nearest 0.05 kg and heights to the nearest 0.1 cm were measured for participants dressed in light examination clothing without shoes.

Study Population

We used data from the Examination Survey's Mexican-American portion on the Mexican-American males and nonpregnant females who were 12 through 19 years old. Analyses were done on the 392 males and 437 females who self-responded, reported weights and heights,

At the time of the study the authors were with the Division of Health Examination Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Md.

Requests for reprints should be sent to Harold Davis, MD, US AID, 1111 N 19th St, Suite 300, Rosslyn, VA 22209.

This paper was accepted June 3, 1993.

TABLE 1—Mean Percentage Difference between Reported and Measured Values for Weight, Height, and Body Mass Index (BMI): Mexican Americans, 12 through 19 Years Old, Hispanic Health and Nutrition Examination Survey, 1982 through 1984

Group	Percentage Difference for Weight (Mean ± SEM)		Percentage Difference for Height (Mean ± SEM)		Percentage Difference for BMI (Mean ± SEM)	
	Males	Females	Males	Females	Males	Females
Measured BMI group						
Low	4.6 ± 0.9[a]	2.4 ± 0.6[a]	−0.2 ± 0.4	−0.2 ± 0.2[a]	5.1 ± 1.1[a]	2.9 ± 0.6[a]
Middle	1.2 ± 0.5	−1.7 ± 0.3	−0.2 ± 0.2	0.4 ± 0.1	1.8 ± 0.6	−2.2 ± 0.4
High	−4.1 ± 1.0	−6.9 ± 1.1	<0.1 ± 0.4	1.1 ± 0.3	−4.1 ± 1.0[b]	−8.8 ± 1.3
Self-described weight status						
Underweight	2.4 ± 0.9	1.5 ± 0.8[c]	−0.1 ± 0.2	1.2 ± 0.4	0.3 ± 1.2	−0.6 ± 1.2
About right	0.7 ± 0.5	−0.6 ± 0.4	−0.2 ± 0.2	0.2 ± 0.2	1.4 ± 0.6	−0.7 ± 0.6
Overweight	−0.2 ± 1.2	−3.4 ± 0.5	0.1 ± 0.3	0.5 ± 0.1	−0.5 ± 0.1	−4.3 ± 0.6[d]
Age group, y						
12–14	<0.1 ± 1.0	−2.7 ± 0.6	−0.7 ± 0.3[e]	−0.2 ± 0.3	1.6 ± 1.0	−1.9 ± 0.9
15–17	0.8 ± 0.5	−1.7 ± 0.5	−0.1 ± 0.2	−0.7 ± 0.1	1.2 ± 0.7	−3.0 ± 0.6
18–19	1.6 ± 0.6	−1.1 ± 0.4	0.5 ± 0.3	0.5 ± 0.2	0.9 ± 0.9	−2.0 ± 0.6

[a]Measured low body mass index group is significantly different from measured high body mass index group.
[b]Males with measured high body mass index are significantly different from females with measured high body mass index.
[c]Underweight group is significantly different from overweight group.
[d]Overweight group is significantly different from about-right-weight group.
[e]Adolescents 12 through 14 years old are significantly different from those 18 and 19 years old.

and had their weights and heights measured.

Statistical Analyses

We used measured values to calculate body mass index as weight/square of height (kilograms/square meter). Based on the 1308 measured body mass indexes of Mexican-American adolescents in the Examination Survey, single year of age-and-sex-specific cutoffs were created (by using sample weights) for three measured body mass index groups: low (≤15th percentile), middle (10th–84th percentile), and high (≥85th percentile). Reported body mass index was calculated by using reported weights and heights and categorized by using these cutoffs.

Sensitivity and specificity of reported low body mass indexes were calculated as (1) sensitivity = 100 × number of adolescents with reported and measured low body mass index/number with measured low body mass index; (2) specificity = 100 × number of adolescents with neither reported nor measured low body mass index/number without measured low body mass index. Analogous methods were used for reported high body mass index.

For each adolescent, a percentage difference between reported and measured values was calculated as 100 × (reported value − measured value)/measured value. The mean percentage difference is the average of these individual percentage differences.

Point estimates were calculated with sample weights. Sampling weights account for selection probabilities and include adjustments for nonresponse and geographic noncoverage. Variances were estimated by multiplying simple-random-sample variance estimates by the average design effect (1.00 in this study).[21] To compare point estimates, *t* tests with 8 *df* were used. To assess trends, linear regression analysis was performed. *P* values of less than .05 were significant.

Results

Self-Reported Weight

The mean percentage difference between reported and measured weights was 0.8% ± 0.4% (SD = 8.5) for males and −1.8% ± 0.3% (SD = 6.7) for females. Weights were reported within ±5% of measured values by 63% of males and 72% of females. For both sexes, adolescents with measured low body mass index on the average overestimated their weights, and those with measured high body mass index underestimated theirs (Table 1).

The correlation coefficient between measured and reported weight was .95 for males and .93 for females and varied little with age group.

Self-Reported Height

The overall mean percentage difference between reported and measured heights was −0.1% ± 0.2% (SD = 3.2) for males and 0.4% ± 0.1% (SD = 2.4) for females. The reported height fell within ±2% of the measured height for 63% of males and 70% of females. Adolescents 12 through 14 years old on the average slightly underestimated their heights, whereas 18- and 19-year-olds slightly overestimated theirs (Table 1).

The correlation coefficient between measured and self-reported height was .86 for males and .86 for females. This coefficient varied little with age group except for males 18 and 19 years old, for whom it was .71.

Self-Reported Body Mass Index

The overall mean percentage difference between the self-reported and measured body mass indexes was 1.3% ± 0.5% (SD = 9.8) for males and −2.4% ± 0.4% (SD = 8.4) for females. On average, reported weights and heights overestimated the body mass index of adolescents with measured low body mass indexes (Table 1). Reported values tended to underestimate the body mass index of adolescents with measured high

TABLE 2—Cross-Tabulation of Self-Reported and Measured Body Mass Index (BMI) Groups, by Sex: Mexican Americans, 12 through 19 Years Old, Hispanic Health and Nutrition Examination Survey, 1982 through 1984

BMI Groups Based on Self-Reported Weight and Height	BMI Groups Based on Measured Weight and Height, %		
	Low	Middle	High
Males			
Low	7.1	4.1	0.2
Middle	7.6	60.7	3.9
High	0.0	3.7	12.8
Females			
Low	10.4	7.2	0.0
Middle	4.2	63.8	5.5
High	0.0	1.0	7.9

Note. Percentages do not add to 100 because of rounding.

TABLE 3—Sensitivity and Specificity of Low Reported and High Reported Body Mass Index (BMI) Groups: Mexican Americans, 12 through 19 Years Old, Hispanic Health and Nutrition Examination Survey, 1982 through 1984

Group	Sensitivity		Specificity	
	%	95% Confidence Interval	%	95% Confidence Interval
Low reported BMI				
Males	48	33, 63	95	92, 98
Females	71	58, 85	92	88, 95
High reported BMI				
Males	76	63, 88	96	93, 98
Females	59	44, 74	99	98, 100

body mass indexes, with greater underestimation for females than males. On average, reported body mass indexes were underestimates to a greater degree for females self-described as overweight than for those described as underweight or about the right weight.

The correlation coefficient between measured and reported body mass indexes was .87 for males and .85 for females and varied little with age group.

The distribution of reported vs measured body mass index groups is shown in Table 2. In neither sex did the sensitivities for reported low and high body mass index reach 80% (Table 3). The specificities for reported low and high body mass index ranged from 92% to 99%. For each sex, there were no significant trends by age group in sensitivity or specificity.

Discussion

We found that categories of reported body mass indexes had low sensitivities. That indicates that cutoffs based on absolute values had a poor ability to correctly classify adolescents. Therefore, using these categories to estimate relative risks associated with categories of body mass index might result in large errors in the estimates, and it is probably best not to use such categories to estimate relative risks.

We found high correlations between reported and measured values for weight, height, and body mass index, indicating that reported values reflect well the relative ranking of measured values. In multivariate analyses with continuous data, using reported values that are highly correlated with measured values is likely to cause relatively small errors in a variable's coefficient.[22] Therefore, using adolescents' reported values in these analyses might result in only small errors in the coefficients for weight, height, and body mass index.

Similarities between this study and most previous studies of reported values in adults and adolescents include a high correlation between reported and measured values[2,4,7,9,10,14,18,19] and a tendency for persons with high body mass indexes to underestimate their weights more than other persons,[2-10,15-18] a tendency that causes a low sensitivity of high body mass index categories formed from reported values.[7,10,13] These similarities suggest that the limitations and possible uses of reported values of Mexican-American adolescents described here will apply to most other adolescent populations. □

References

1. Schlichting PF, Hoilund-Carlsen PF, Quaade F. Comparison of self-reported height and weight with controlled height and weight in women and men. *Int J Obes.* 1981;5:67–76.
2. Stunkard AJ, Albaum JM. The accuracy of self-reported weights. *Am J Clin Nutr.* 1981;34:1593–1599.
3. Pirie P, Jacobs D, Jeffery R, Hannan P. Distortion in self-reported height and weight data. *J Am Diet Assoc.* 1981;78:601–606.
4. Stewart AL. The reliability and validity of self-reported weight and height. *J Chronic Dis.* 1982;35:295–309.
5. Palta M, Prineas RJ, Berman R, Hannan P. Comparison of self-reported and measured height and weight. *Am J Epidemiol.* 1982;115:223–230.
6. Jalkanen L, Tuomilehto J, Tanskanen A, Puska P. Accuracy of self-reported body weight compared to measured body weight. *Scand J Soc Med.* 1987;15:191–198.
7. Stewart AW, Jackson RT, Ford MA, Beaglehole R. Underestimation of relative weight by use of self-reported height and weight. *Am J Epidemiol.* 1987;125:122–126.
8. Rowland M. Reporting bias in height and weight data. *Stat Bull Metropol Insur Co.* 1989;70(Apr–Jun):2–11.
9. Rowland ML. Self-reported weight and height. *Am J Clin Nutr.* 1990;52:1125–1133.
10. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: sensitivity and specificity using self-reported weight and height. *Epidemiology.* 1990;1:146–152.
11. Millar WJ. Distribution of body weight and height: comparison of estimates based on self-reported and observed measures. *J Epidemiol Community Health.* 1986;40:319–323.
12. Kuskowska-Wolk A, Karlsson P, Stolt M, Rossner S. The predictive validity of body mass index based on self-reported weight and height. *Int J Obes.* 1989;13:441–453.
13. Kuskowska-Wolk A, Rossner S. Influence of body image on estimation of body mass index based on self-reported weight and height. *Diabetes Res Clin Pract.* 1990;10:S155–S158.
14. Wing RR, Epstein LH, Ossip DJ, LaPorte RE. Reliability and validity of self-report and observers' estimates of relative weight. *Addict Behav.* 1979;4:133–140.
15. Dwyer JT, Feldman JJ, Seltzer CC, Mayer J. Adolescent attitudes toward weight and appearance. *J Nutr Educ.* 1969;1:14–19.
16. Shannon B, Smickilas-Wright H, Wang MQ. Inaccuracies in self-reported weights and heights of a sample of sixth-grade children. *J Am Diet Assoc.* 1991;91:675–678.

17. Fortenberry JD. Reliability of adolescents' reports of height and weight. *J Adolesc Health.* 1992;13:114–117.

18. Himes JH, Story M. Validity of self-reported weight and stature of American Indian youth. *J Adolesc Health.* 1992;13: 118–120.

19. Tienboon P, Wahlqvist ML, Rutishauser IHE. Self-reported weight and height in adolescents and their parents. *J Adolesc Health.* 1992;13:528–532.

20. National Center for Health Statistics. Plan and operation of the Hispanic Health and Nutrition Examination Survey 1982–1984. *Vital Health Stat* [1]. September 1985; no. 19. DHHS publication PHS 85-1321.

21. Kovar MG. Approaches for the analysis of data. *Vital Health Stat* [1]. 1985; no. 19:56–67. DHHS publication PHS 85-1321.

22. Keyl PM, Flegal KM, Nieto-Garcia FJ. Effects of using self-reported versus measured weight and height in epidemiologic analyses. *Am J Epidemiol.* 1991;134:733–734.

The Prevalence of Intrauterine Growth Retardation in Mexican Americans

Hector Balcazar, PhD, MS

ABSTRACT

This study evaluated the prevalence of intrauterine growth retardation in Mexican Americans compared with non-Hispanic Whites in the state of Arizona. Data came from all live birth certificates in 1986 and 1987. Rates of intrauterine growth retardation in Mexican-American and non-Hispanic White infants in Arizona were lower than those in White infants in California. Differences in patterns of the 10th percentile growth distribution curves were observed between infants born in Arizona and those born in California. Compared with non-Hispanic Whites, Mexican Americans had lower adjusted odds ratios for intrauterine growth retardation according to several maternal risk characteristics. (*Am J Public Health.* 1994; 84:462–465)

Introduction

Given the heterogeneity of the Mexican-American population, the study of perinatal health problems should not be confined to low birthweight.[1] In the Mexican-American perinatal population, indicators of fetal growth and maturity at birth have not been used to evaluate intrauterine growth retardation. It is quite likely, therefore, that reported prevalence of low birthweight overlooks a proportion of affected infants. Failure to diagnose intrauterine growth retardation at birth in Mexican-American infants, whether or not they fall into the low-birthweight category, has public health significance.[1,2] Evidence is beginning to accumulate in regard to the detrimental effects on morbidity and mortality in newborns with birthweights above 2500 g.[3–5] This study was designed to evaluate the problem of intrauterine growth retardation in a Mexican-American population.

Methods

All singleton births in Arizona for the years of 1986 and 1987 were evaluated for intrauterine growth retardation. The entry of Spanish origin and race of the mother was used to identify Whites versus Hispanics and those of Mexican origin. The total numbers of single births were 23 799 and 67 280 for Mexican-Americans and Whites, respectively. The percentages of births that occurred out of the hospital infirmary or medical center were 1.9% for Mexican Americans and 2.0% for Whites. Approximately 1.3% of the total birth certificates for each ethnic group were discarded because of missing data on birthweight. Similar percentages of missing data in regard to origin or race of the mother were also found (1% of the total population). For Mexican Americans and Whites, 3.9% and 2.5% of the data, respectively, were missing for the month of the last menstrual period. The prevalence of intrauterine growth retardation was estimated for all infants with birthweight data at 24 to 42 weeks of gestation.

Completed weeks of gestation were computed based on the last normal menses and the date of birth of the infant. If the day of the last normal menses was missing, the 15th day of that month was used. For the purpose of classifying intrauterine growth retardation, the cutoff point of the 10th percentile of each of two different fetal growth distributions was used (Williams et al.[6] and Hoffman et al.[7]). The White (non-Hispanic) population of Arizona was used as a comparison group.

Several maternal risk factors were defined from the certificates of live birth to evaluate the variations in classification of intrauterine growth retardation. Chi-square statistics were used to test for differences in prevalences between groups. Adjusted odds ratios and 95% confidence intervals were calculated to

Hector Balcazar is with the Department of Family Resources and Human Development, Arizona State University, Tempe, Ariz.

Requests for reprints should be sent to Hector Balcazar, PhD, MS, Community and International Nutrition, Department of Family Resources and Human Development, Arizona State University, Tempe, AZ 85287-2502.

This paper was accepted June 28, 1993.

TABLE 1—Distribution of Newborn Infants, According to Low-Birthweight Classification and Gestational Age, between Mexican Americans and Non-Hispanic Whites

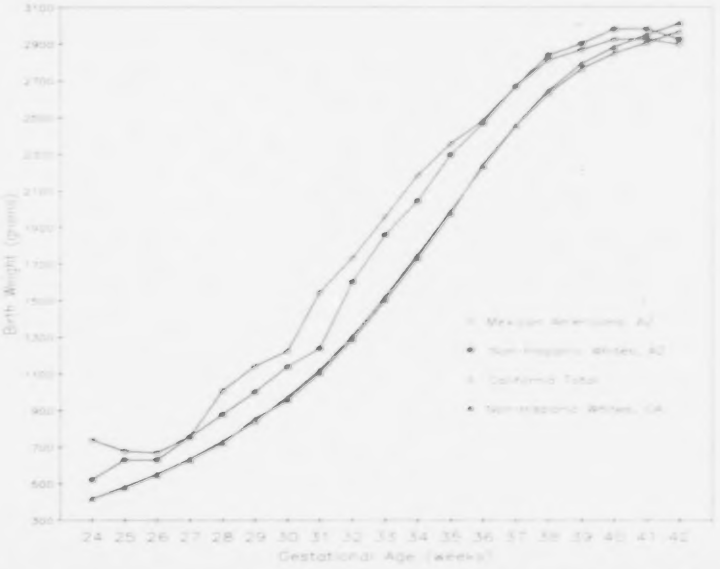
Newborn Classification	Mexican Americans (n = 23 799), %	Whites (n = 67 280), %
< 37 weeks of gestation	14.2	11.1*
< 2500 g	5.1	5.0
< 37 weeks of gestation; < 2500 g	3.1	3.0
≥ 37 weeks of gestation; < 2500 g	2.0	2.0

Note. For the total sample, gestational ages were between 24 and 42 weeks.
*$P < .01$ by chi-square test.

TABLE 2—Prevalence of Intrauterine Growth Retardation in Mexican Americans and Non-Hispanic Whites in Arizona Compared with Two Reference Fetal Growth Distributions

Reference Fetal Growth Distribution	Mexican Americans (n = 23 799), %	Whites (n = 67 280), %	P*
California Non-Hispanic Growth Data[6]	6.3	5.8	.015
National Growth Data[7]	4.9	4.7	.215

Note. Sex-specific cutoff values were used for 10th percentile of each reference growth standard. For the total sample, gestational ages were between 24 and 42 weeks.
*P value for differences in prevalence between Mexican Americans and Whites for the two reference fetal growth distributions according to the chi-square test.



Note. Sample size ranges for Mexican Americans in Arizona are 19–1339 (preterm) and 1219–5296 (term); for non-Hispanic Whites in Arizona they are 50–3244 (preterm) and 3102–16 192 (term). Tenth percentile birthweight values for each gestational age group in the Arizona samples are available from the author on request.
Source. Data for California samples is from Table 2 of Williams et al.[6] and is used with permission from The American College of Obstetricians and Gynecologists.

FIGURE 1—Comparison of 10th percentile cutoff values (birthweight by gestational age) reported for Mexican Americans and Whites in Arizona and California.

evaluate the statistical significance of differences in intrauterine growth retardation between ethnic groups within each level of maternal risk factors. Multivariate logistic regression was used to evaluate differences in intrauterine growth retardation for each maternal risk factor after controlling for the effects of the other maternal characteristics. Data were analyzed on an IBM mainframe computer with the SAS statistical package.[8]

Results

Mexican Americans have a 3.1% excess of preterm delivery compared with Whites (Table 1). Comparable rates of low birthweight are observed for Mexican Americans and Whites. Similar results are also seen for preterm and term low-birthweight groups. The proportions of low birthweight associated with preterm delivery are similar for Mexican Americans and Whites (60.8% vs 60.0%). Among preterm infants, the prevalence of low birthweight differs by ethnic group: 21.6% for Mexican Americans vs 27.2% for Whites ($P < 0.01$).

The prevalence of intrauterine growth retardation for Mexican Americans is 4.9% and 6.3%, respectively, for each of the two reference growth distributions (Table 2). There was a significant difference in prevalence of intrauterine growth retardation between Mexican Americans and Whites (6.3% vs 5.8%, respectively) only with the fetal growth distribution in California.[6] The California reference distribution indicates that more than 60% of the Mexican-American and White infants with intrauterine growth retardation had birthweights greater than or equal to 2500 g. These results are striking considering the fact that 942 Mexican-American and 2398 White infants with intrauterine growth retardation would not have been identified by the low-birthweight criterion alone. Similarly, the two reference growth distributions classify more than 85% of all case subjects with intrauterine growth retardation as term infants in both Mexican Americans and Whites (results not shown).

Figure 1 clearly illustrates the different patterns of birthweight distribution for the 10th percentile cutoff values in

TABLE 3—Intrauterine Growth Retardation[a] among Mexican-American and Non-Hispanic White Infants, by Maternal Risk Characteristics

Maternal Risk Factor	Prevalence of Intrauterine Growth Retardation, %		Adjusted Odds Ratio[c]	95% Confidence Interval
	Mexican Americans	Whites[b]		
Education				
Fewer than 9 years	6.22	10.41	0.59	0.51, 0.69
9 or more years	6.29	5.64	0.98	0.92, 1.06
Marital status				
Unmarried	7.79	9.12	0.84	0.76, 0.93
Married	5.38	5.21	0.93	0.86, 1.01
Age				
19 y or younger	8.19	8.17	0.95	0.83, 1.09
20 y or older	5.79	5.57	0.89	0.83, 0.96
Presence of pregnancy complications				
Yes	8.63	7.90	1.01	0.84, 1.21
No	6.05	5.58	0.89	0.83, 0.95
Presence of current illness				
Yes	7.79	7.09	0.92	0.71, 1.18
No	6.19	5.76	0.90	0.85, 0.97
Presence of labor complications				
Yes	7.44	6.85	0.93	0.84, 1.02
No	5.58	5.18	0.89	0.82, 0.97
Lack prenatal care				
Yes	9.42	13.13	0.64	0.46, 0.89
No	6.14	5.76	0.92	0.86, 0.98

[a]Based on gestational ages between 24 and 42 weeks and the 10th percentile of sex-specific birthweight by gestational age distribution used by Williams et al.[6]
[b]Differences in intrauterine growth retardation within ethnic groups for each variable are all statistically significant ($P < .05$) except for education in Mexican Americans.
[c]The odds ratio was calculated by using the maternal risk characteristic distribution of non-Hispanic Whites. These odds ratios were adjusted for all the other maternal risk characteristics.

Arizona vs California populations of infants. The cutoff birthweight values for the Arizona population (including both Mexican Americans and Whites) are higher than those for the California population for each week of gestational age (excepting the 42nd week). Differences in the distribution of birthweights between Mexican Americans and Whites in Arizona are also noticeable. Note that for the Mexican-American subgroup (as well as for the White sample) the 10th percentile cutoff for term infants exceeds the 2500-g low-birthweight criterion.

Differences in intrauterine growth retardation according to maternal characteristics within each ethnic group are all statistically significant, except for education in Mexican Americans (Table 3). In addition, significantly lower adjusted odds ratios associated with intrauterine growth retardation were found in Mexican Americans compared with Whites for at least one level within each of the maternal characteristics shown in Table 3.

Discussion

This analysis clearly demonstrates that the classification of low birthweight (birthweight less than 2500 g) fails to identify infants with intrauterine growth retardation or those who are small for gestational age. Even though there were differences in intrauterine growth retardation between Mexican Americans and Whites, the rates of intrauterine growth retardation for both groups were considerably lower than the 10% rate expected based on that for non-Hispanic Whites in the California reference growth distribution (10th percentile).

Three major factors that must be acknowledged when exploring the prevalence of intrauterine growth retardation in Mexican Americans are the reference fetal growth distribution used,[9] the evaluation of the effect of measurement error associated with estimation of gestational age,[10,11] and the presence of sampling bias or systematic error in determining birthweight or gestational age.

Differences in prevalences of intrauterine growth retardation between groups are probably related to the higher birthweight values observed for Mexican Americans and Whites in Arizona throughout the range of gestational ages compared with the California data set. Factors that might have shifted the distribution towards higher birthweights in Mexican Americans and Whites in Arizona are related to underestimation of gestational age (particularly for preterm or near-term infants) and the exclusion of a sample of infants because of unknown birthweight, unknown origin or race of the mother, or unknown gestational age. The overall effect of these errors probably resulted in underestimation of the prevalence of intrauterine growth retardation in both Mexican Americans and Whites in Arizona. However, it is possible that this phenomenon of higher birthweights observed in Arizona vs California infants is real, rather than artificial.[12]

The large share of intrauterine growth retardation accounted for by term and non–low birthweight infants could have been partially accounted for by measurement error in gestational age estimation. Kramer et al.[13] presented evidence that overestimation of gestational age in postterm gestations due to either delayed ovulation or missed spontaneous abortions may play an important role in explaining systematic errors in measurement of gestational age. In this regard the crossover effect of birthweight observed between the populations of Arizona and California shown in Figure 1 (at 42 weeks of gestation) needs to be carefully examined. Note that the same crossover effect is also observed between Mexican Americans and Whites in Arizona (Figure 1). This effect has also been shown in California.[12] Problems associated with differential underreporting of very small and preterm infants, as suggested by ethnic differences found in preterm infants weighing less than 2500 g (Table 1), need to be examined also.

The rates of intrauterine growth retardation in Mexican Americans described here suggest the need to evaluate the appropriate reference growth distribution to be used for the population of Mexican-American infants in Arizona. Thus, population based–clinical studies are needed to explore this issue (including the evaluation of morbid-

ity and mortality outcomes according to ethnicity and socioeconomic status). In this regard, the use of a criterion less restrictive than the 10th percentile should also be tested when evaluating intrauterine growth retardation in these groups (i.e., the use of standard or percentage deviations from mean birthweight throughout all gestational ages).[14–17] Finally, information on maternal determinants of intrauterine growth retardation between groups was not available for investigation in order to confirm the results of the risk ratios presented. □

Acknowledgments

This research was supported by funds from the Faculty Grant-in-Aid Program, College of Liberal Arts and Sciences, Arizona State University.

I wish to acknowledge the valuable comments on the manuscript provided by Dr Sara Santana from the Department of Public Health, Maricopa Department of Health Services, Phoenix, Ariz.

References

1. Balcazar H, Aoyama C, Cai Xi. Interpretative views on Hispanics' perinatal problems of low birth weight and prenatal care. *Public Health Rep.* 1991;164:420–426.
2. Villar J, Gonzalez T. Nutritional factors associated with low birth weight and short gestational age. *Clin Nutr.* 1986;5:78–85.
3. Haas J, Balcazar H, Caulfiled L. Variations in early neonatal mortality for different types of intrauterine growth retardation. *Am J Phys Anthropol.* 1987;73:467–473.
4. Balcazar H, Haas J. Classification schemes of small-for-gestational age and type of intrauterine growth retardation and its implications to early neonatal mortality. *Early Hum Dev.* 1990;24:219–230.
5. Lester B, Garcia-Coll C, Valcarcel M, Hoffman J, Brazelton TB. Effects of atypical patterns of fetal growth on newborn (NBAS) behavior. *Child Dev.* 1986;57:11–19.
6. Williams R, Creasy R, Cunningham G, et al. Fetal growth and perinatal viability in California. *Obstet Gynecol.* 1982;59:624–632.
7. Hoffman H, Stark C, Lundin F, et al. Analysis of birth weight, gestational age, and fetal viability, U.S. births, 1968. *Obstet Gynecol Surv.* 1974;29:651–681.
8. SAS Institute Inc. *SAS/STAT User's Guide, Version 6.03 Edition.* Cary, NC: SAS Institute Inc; 1988.
9. Goldenberg R, Cutter G, Hoffman H, Foster J, Nelson K, Hauth J. Intrauterine growth retardation: standards for diagnosis. *Am J Obstet Gynecol.* 1989;161:271–277.
10. David R. The quality and completeness of birth weight and gestational age data in computerized birth files. *Am J Public Health.* 1980;70:964–973.
11. Alexander G, Tompkins M, Cornely D. Gestational age reporting and preterm delivery. *Public Health Rep.* 1990;105:267–275.
12. Williams R. Intrauterine growth curves: intra- and international comparisons with different ethnic groups in California. *Prev Med.* 1975;4:163–172.
13. Kramer M, McLean F, Boyd M, Usher R. The validity of gestational age estimation by menstrual dating in term, preterm, and postterm gestations. *JAMA.* 1988;260:3306–3308.
14. Kramer M, McLean F, Olivier M, Willis D, Usher R. Body proportionality and head and length "sparing" in growth-retarded neonates: a critical reappraisal. *Pediatrics.* 1989;84:717–723.
15. Chard T, Costeloe K, Leaf A. Evidence of growth retardation in neonates of apparently normal weight. *Eur J Obstet Gynecol Reprod Biol.* 1992;45:59–62.
16. Fetal growth and birth weight, II: norms and thresholds. In: Kline J, Stein Z, Susser M, eds. *Conception to Birth. Epidemiology of Prenatal Development.* New York: Oxford University Press; 1989:219–230.
17. Balcazar H. Mexican Americans' intrauterine growth retardation and maternal risk factors. *Ethn Dis.* 1993;3:169–175.

DHHS Minority Health Resource Center Established

The US Department of Health and Human Services, Office of Minority Health, has established a resource center to provide health professionals with resources and referrals on minority health. The Office of Minority Health Resource Center operates a toll-free telephone number for inquiries on the health status of Asian, Black, Hispanic, Native American, and Pacific Islander populations.

The center's services and products include the following:

- **The center conducts database searches for documents on minority health, national organizations, and community-based programs.**
- **It supplies copies of research articles and other relevant materials on more than 100 subjects.**
- **It provides technical assistance.**
- **It links users with experts nationwide for conferences and program efforts.**
- **It refers users to data resources.**
- **It provides minority media resources.**
- **It distributes a newsletter, resource lists, grantsmanship materials, and federal health promotion/disease prevention books.**

These services and products are offered free of charge, and the library is accessible to visitors. To obtain information or schedule an appointment, write or call the Office of Minority Health Resource Center, PO Box 37337, Washington, DC 20013-7337; tel 1-800-444-6472.

The Proportion of Veterans among Homeless Men

Robert Rosenheck, MD, Linda Frisman, PhD, and An-Me Chung, PhD

ABSTRACT

This study used data from four surveys conducted in 1986/87 to identify age–race cohorts of homeless men in which veterans are overrepresented. The overall proportion of veterans among homeless men (41%) was somewhat higher than that in the general population (34%). This overrepresentation is largely attributable to the disproportionate representation of veterans in the youngest age cohort (20 to 34 years) of homeless White men. Veterans in this group are 4.76 times more likely to be homeless than nonveterans. Other national data on this cohort reveal higher rates of unemployment, substance abuse, and antisocial personality among veterans than among nonveterans. (*Am J Public Health.* 1994; 84:466–469)

Introduction

Community surveys conducted during the 1980s reported that, on average, 38% of homeless men were veterans (range = 29% to 49%),[1–7] as compared with only 30% of men in the general population.[8] While these findings suggest that veterans may be at a somewhat greater than expected risk for homelessness, such an overrepresentation is not easy to explain. Most veterans are entitled to special benefits, including educational assistance, home loan guarantees, pension and disability payments, and free health care. In 1990, veterans had higher median incomes, had lower rates of poverty and unemployment, and were better educated than nonveteran US men in similar age groups.[9] These data suggest that veterans should be *less* vulnerable to homelessness than other Americans. The present study sought to determine whether veterans, in general, are disproportionately represented among homeless people or whether vulnerability to homelessness is specific to one or another age or race cohort of veterans.

Methods

Estimates of Proportions

The proportion of male veterans among homeless men was estimated through secondary analyses of the Urban Institute's 1987 national survey of homeless service users (n = 1140)[10] and three single-city surveys conducted in Los Angeles (n = 270),[11] Baltimore (n = 295),[12] and Chicago (n = 486)[13] between 1985 and 1987. These studies were chosen because of their sophisticated sampling strategies (see the original publications for details) and overall rigorous design. Veteran status, judged in all four studies from a single question concerning past US military service, was used to estimate the proportion of veterans among homeless men within each of five age cohorts (20 through 34, 35 through 44, 45 through 54, 55 through 64, >64) and by race (White vs Black).

The 1987 Current Population Survey, a national household survey conducted by the Bureau of the Census, was used to estimate the proportion of male veterans in the general population in each of these age–race cohorts.[14] These data were used to estimate odds ratios (ORs) of homelessness among veterans as compared with nonveteran US men.

Adjustment Factors Associated with Veteran Status

To account for differences in the risk of homelessness among veterans relative to nonveterans in the various age–race cohorts, we also compared veterans and nonveterans in the general population on factors that have been associated with homelessness.

Socioeconomic characteristics. Further analyses of the Current Population Survey were undertaken to compare veterans and nonveterans on personal income, poverty rates, unemployment, educational achievement, marital status, and household size.[14] Only the subgroup of men whose income was less than twice the poverty level were included in the last four analyses.

Psychiatric and substance abuse disorders. An analysis of the public access data tape of the Epidemiological Catchment Area Study was conducted to compare rates of lifetime psychiatric and substance abuse diagnoses among male veterans and nonveterans in each 1987 age–race cohort.[15]

Military service. Information regarding service era and combat exposure of veterans in various age groups was obtained through analyses of data from the Third Survey of Veterans, a 1987 survey of a representative national

The authors are with the Department of Veterans Affairs Medical Center, Northeast Program Evaluation Center, West Haven, Conn. Robert Rosenheck and Linda Frisman are also with the Department of Psychiatry, Yale University, New Haven, Conn. An-Me Chung is also with the College of Physicians and Surgeons, Columbia University, New York, NY.

Requests for reprints should be sent to Robert Rosenheck, MD/182, Department of Veterans Affairs Medical Center, Northeast Program Evaluation Center, 950 Campbell Ave, West Haven, CT 06516.

This paper was accepted July 13, 1993.

sample of 9442 noninstitutionalized veterans.[16] These analyses were also limited to low-income veterans: those who earned less than $15 000 per year (the bottom third of the veteran income distribution).

Results

Veterans among Homeless Americans

Results were similar from all four homeless surveys, and unweighted data were combined to simplify their presentation. Among homeless men more than 20 years old, 41% reported past military service, as compared with only 34% in the general population of US men surveyed in the 1987 Current Population Survey (Table 1).

Odds ratios (Table 2) show significantly greater likelihood of homelessness among veterans 20 to 34 and 45 to 54 years of age than among nonveterans in the general population. The most dramatic finding is that White veterans 20 to 34 years old were 4.8 times more likely to be homeless than their nonveteran counterparts in the general population (95% confidence interval [CI] = 3.8, 5.9). Although the risk of homelessness was also elevated among Black veterans in this age group, the odds ratio (2.1, 95% CI = 1.7, 2.7) was significantly lower than that for Whites. Homelessness was also increased among veterans 45 to 54 years old (OR = 1.7, 95% CI = 1.4, 2.1), but differences between races were not significant.

TABLE 1—Veterans among Homeless Men and among Men in the General Population, by Age and Race

Age, y	White		Black		Total Sample[a]	
	No.	% Veterans	No.	% Veterans	No.	% Veterans
Homeless surveys (n = 2223)						
20–34	400	33.7	483	25.6	939	30.6
35–44	253	42.4	297	33.6	576	37.2
45–54	240	64.6	161	48.0	412	58.7
55–64	154	67.4	89	51.9	251	61.7
>64	22	67.8	23	7.1	45	37.4
Total	1 071	48.2	1 052	33.1	2 223	41.2
1987 Current Population Survey (thousands)						
20–34	25 522	9.6	3 487	13.8	30 021	10.0
35–44	14 202	38.0	1 566	33.3	16 310	36.9
45–54	9 696	49.9	1 805	21.9	11 845	44.8
55–64	9 133	72.2	941	53.9	10 304	69.9
>64	10 418	47.5	941	37.6	11 550	46.3
Total	68 971	35.1	8 740	25.9	80 029	33.6

[a]Includes Whites, Blacks, and others.

TABLE 2—Odds Ratios of Homelessness among Veterans Compared with Nonveterans

Age, y	White		Black		Total Sample[a]	
	OR	95% CI	OR	95% CI	OR	95% CI
20–34	4.76	3.83, 5.93	2.13	1.68, 2.72	3.95	3.39, 4.58
35–44	1.19	0.92, 1.55	0.95	0.66, 1.37	1.01	0.85, 1.21
45–54	1.83	1.39, 2.42	3.25	2.29, 4.63	1.75	1.45, 2.15
55–64	1.42	0.56, 1.14	0.91	0.57, 1.45	0.69	0.53, 0.91
>64	2.36	0.91, 6.40	0.15	0.04, 0.70	0.71	0.37, 1.34
Total	1.71	1.52, 1.94	1.42	1.22, 1.63	1.38	1.05, 1.85

Note. OR = odds ratio; CI = confidence interval.
[a]Includes Whites, Blacks, and others.

Socioeconomic Factors

Data from the 1987 Current Population Survey show that the average personal income among White veterans 20 to 34 years of age was about the same as that of White nonveterans ($17 490 vs $17 653), as were poverty rates (10.0% vs 10.8%). White veterans aged 20 to 34 were more likely to be unemployed than nonveterans (18.6% vs 11.4%), but veterans of all ages and races, including White veterans aged 20 to 34, were more likely to have a high school degree, more likely to be married, and no more likely to be living alone than their nonveteran counterparts (data available on request).

Psychiatric Risk Factors

Among White men in the 20- to 34-year-old cohort, prevalences of non–substance abuse psychiatric disorders and substance abuse disorders were two to three times higher among veterans than among nonveterans, and rates of antisocial personality disorder were five to six times as high (Table 3). In contrast, Black veterans in this age cohort, in comparison with Black nonveterans, showed a significantly greater prevalence for lifetime substance abuse disorder only. Other veteran age–race groups, for the most part, showed no greater rates of these disorders than nonveterans.

Military Service Characteristics

Data from the Third Survey of Veterans revealed that veterans in the 20 to 34 and 45 to 54 cohorts were far less likely than other veterans to have served during a wartime service era or to have been exposed to combat (data available on request). Among the 20 to 34 age group, 79% served during the post-Vietnam era (the era of the All Volunteer Force), and only 7% served in combat. Among the 45 to 54 age group, 78% served between the Korean and Vietnam eras, and only 17% served in combat. More than 40% of veterans in other age groups, in contrast, reported combat service.

Discussion

The proportion of veterans among homeless male Americans is greater than the proportion in the general population, but this overrepresentation is most striking among Whites 20 to 34 years old; to a lesser extent, Blacks 20 to

TABLE 3—Lifetime Prevalence (%) of Psychiatric Disorder among Men in the Epidemiological Catchment Area Study of Five US Cities, by 1987 Age Cohort, Race, and Veteran Status

		Non–Substance Abuse		Substance Abuse		Antisocial Personality	
Race/Age, y	n	Veterans	Non-veterans	Veterans	Non-veterans	Veterans	Non-veterans
White							
20–34	1146	39.0	16.5***	53.3	30.2***	24.7	4.8***
35–44	1255	21.0	22.7	38.6	31.1**	9.5	7.2
45–54	663	15.4	16.3	25.4	27.3	1.6	4.5**
55–64	581	13.1	20.6**	20.2	22.7	4.3	3.2
>64	2284	11.7	12.2	14.6	13.2	0.7	0.5
Black							
20–34	346	12.0	25.4	33.6	15.0*	4.9	2.2
35–44	324	21.0	22.2	39.5	22.0*	8.3	3.5
45–54	166	14.2	25.5	27.2	42.7	3.7	4.6
55–64	141	19.8	24.6	29.2	34.6	3.5	0.8
>64	451	20.7	34.0	28.0	24.7	0.8	1.6

*$P < .10$, **$P < .05$; ***$P < .001$.

34 years of age and 45- to 54-year-old men of both races are overrepresented. Although Current Population Survey data are national rather than strictly urban, as in the case of the survey data, they suggest no special socioeconomic vulnerabilities among high-risk age–race cohorts of veterans. The much higher prevalence of psychiatric illness, substance abuse, and, especially, antisocial personality disorder among White veterans 20 to 34 years old would seem to be a likely contributor to the greater vulnerability to homelessness of this group. Veterans in these most vulnerable groups are the least likely to have served during wartime or in combat, reducing the role that can be attributed to combat stress in the genesis of homelessness among veterans.

Perhaps the most distinctive characteristic of veterans in the vulnerable age cohorts is that they served during nonwartime eras, especially during the post-Vietnam era, after establishment of the All Volunteer Force in 1973. The shift to the volunteer force marked an important change in military manpower policy, a change that was accompanied by widespread concern that the military would become an employer of last resort for poorly skilled youth.[17,18] Studies comparing volunteer force recruits and age-matched civilians show these recruits to have had slightly lower socioeconomic status than their nonveteran peers,[19–21] as well as poorer intellectual aptitude test results[22,23] and greater problems with substance abuse.[24–26]

Although it appears that, during the post-Vietnam era, young adults with greater psychopathology and lower earning potential joined the military, this does not appear to have been the case during the period between Korea and Vietnam.[17] Nonwartime era veterans, however, are eligible for a more limited range of Veterans Affairs benefits. Only 25% of post–Korean era veterans used GI Bill educational benefits, as compared with 54% of Vietnam era veterans.[16] More limited availability of Veterans Affairs benefits may therefore have played some role in increasing the risk of homelessness among veterans in the 45 to 54 age cohort.

Since veteran status in this study was not validated through a review of military records, it is possible that, either as a point of pride or in the hope of obtaining some other advantage, some homeless respondents may have falsely claimed military service, thus inflating estimates of the number of veterans among homeless people. This seems unlikely, however, since there is no reason to expect such falsification in some age–race cohorts more than in others.

Since the homeless surveys involved prevalence rather than incidence samples, an association between veteran status and duration of homelessness could have biased our results.[27] Other studies, however, suggest no difference between veterans and nonveterans in duration of homelessness.[28]

Veterans appear to be overrepresented among homeless men, but this overrepresentation primarily affects younger veterans who served in nonwartime eras, especially the post-Vietnam era. The increased vulnerability for homelessness among veterans may be related to the admission of poorly adjusted young men to military service during nonwartime eras and to the reduced availability of Veterans Affairs benefits to these veterans. □

Acknowledgments

We gratefully acknowledge the generosity of the researchers who shared their survey data with us: William Breakey, MD, Martha Burt, PhD, Paul Koegel, PhD, Philip Leaf, PhD, and Peter Rossi, PhD. Alan Fontana, PhD, Catherine Leda, MSN, MPH, Ellen Fisher, PhD, Peggy Gallup, PhD, and Eric Lindblom provided valuable comments on earlier drafts. Ezra Susser, MD, PhD, provided valuable consultation. Dennis Thompson, Louis Massari, Dee Weber, and Louis Florio assisted with various aspects of data management and computer programming.

References

1. Schutt RK. *Boston's Homeless: Their Backgrounds, Problems, and Needs.* Boston, Mass: University of Massachusetts; 1985.
2. Gelberg L, Linn LS, Leake J. Mental health, alcohol and drug use and criminal history among homeless adults. *Am J Psychiatry.* 1988;145:191–196.
3. Robertson MJ. *The Homeless in Los Angeles County: An Empirical Assessment.* Los Angeles, Calif: UCLA School of Public Health; 1985.
4. Crystal S. *Chronic and Situational Dependency: Long Term Residents in a Shelter for Men.* New York, NY: City of New York Human Resources Administration; 1982.
5. Streuning EL, Rosenblatt A. Characteristics of homeless veterans in the New York City shelter system. New York, NY: New York State Psychiatric Institute; 1987. Working paper.
6. Morse G, Shields NM, Hanneke CR, Calsyn RJ, Burger GK, Nelson B. *Homeless People in St. Louis.* Jefferson City, Mo: Missouri State Department of Mental Health; 1985.
7. Wright J, Weber E. *Homelessness and Health.* Washington, DC: McGraw-Hill; 1987.
8. *Statistical Abstracts of the United States: 1990.* 110th ed. Washington, DC: US Bureau of the Census; 1990.
9. *Annual Report, 1990.* Washington, DC: Department of Veterans Affairs; 1990.
10. Burt MR, Cohen BE. *Feeding the Homeless: Does the Prepared Meals Provision Help?* Washington, DC: The Urban Institute; 1988.
11. Koegel P, Burnam A, Farr R. The prevalence of specific psychiatric disorders among homeless individuals in the

inner-city of Los Angeles. *Arch Gen Psychiatry.* 1989;45:1085–1092.
12. Breakey W, Fischer P, Kramer M, et al. Health and mental health problems of homeless men and women in Baltimore. *JAMA.* 1989;262:1352–1361.
13. Rossi PH, Fisher GA, Willis G. *The Condition of the Homeless in Chicago.* Amherst, Mass: Social and Demographic Research Institute; 1986.
14. *Current Population Survey: Annual Demographic File, 1987.* Washington, DC: US Dept of Commerce; 1987.
15. Leaf PJ, Myers JK, McEvoy LT. Procedures used in the Epidemiologic Catchment Area Study. In: Robins LN, Regier DA, eds. *Psychiatric Disorders in America.* New York, NY: Free Press; 1991.
16. US Bureau of the Census. *1987 Survey of Veterans.* Washington, DC: Department of Veterans Affairs; 1989.
17. Berryman SE. *Who Serves? The Persistent Myth of the Underclass Army.* Boulder, Colo: Westview Press; 1988.
18. Janowitz M. The all-volunteer military as a socio-political problem. *Soc Problems.* 1975;22:432–449.
19. Cooper RVL. *Military Manpower and the All-Volunteer Force.* Santa Monica, Calif: Rand; 1977.
20. Kim C, Nestel G, Phillips RL, Borus ME. *National Longitudinal Survey of Youth Labor Market Experience: Military Studies.* Columbus, Ohio: Center for Human Resources Research, Ohio State University; 1980.
21. *Population Representation in the Military Service: Fiscal Year 1989.* Washington, DC: Office of the Assistant Secretary of Defense; 1990.
22. Eitelberg MJ, Laurence JH, Waters BK, Perelman LS. *Screening for Service: Aptitude and Education Criteria for Military Entry.* Washington, DC: Office of the Assistant Secretary of Defense; 1984.
23. Laurence JH, Ramsberger PF, Gribben MA. *Effects of Military Experience on the Post-Service Lives of Low-Aptitude Recruits: Project 100,000 and the ASVAB Misnorming.* Alexandria, Va: Human Resources Research Organization; 1989.
24. Cahalan D, Cisin IA, Gardner GL, Smith GC. *Drinking Practices and Problems in the U.S. Army, 1972.* Washington, DC: US Dept of the Army; 1972.
25. Bray RM, Marsden ME, Guess LL, et al. *1985 Worldwide Survey of Alcohol and Nonmedical Drug Use among Military Personnel.* Research Triangle Park, NC: Research Triangle Institute; 1986.
26. Polich JM. Epidemiology of alcohol abuse in military and civilian populations. *Am J Public Health.* 1981;71:1125–1132.
27. Susser E, Streuning E, Conover S. Psychiatric problems of homeless men. *Arch Gen Psychiatry.* 1989;46:845–850.
28. Rosenheck RA, Leda CL, Gallup P, et al. Initial assessment data from a 43-site program for homeless chronically mentally ill veterans. *Hosp Community Psychiatry.* 1989;40:937–942.

The Driver's License List as a Population-Based Sampling Frame in Iowa

Charles F. Lynch, MD, PhD, Nyla Logsden-Sackett, MS, Sandra L. Edwards, MA, and Kenneth P. Cantor, PhD, MPH

ABSTRACT

Driver's license lists are infrequently used for population-based sampling, presumably because of suspicions of poor population coverage. The 1990 Iowa driver's license list was compared with the 1990 census to evaluate coverage by 5-year age group, sex, resident county, and urbanicity. Coverage exceeded 90% among 15- to 74-year-old men and 15- to 64-year-old women, with uniform coverage by county and county urbanicity group in these age ranges. In Iowa, these lists are convenient and cost-effective and appear to be representative for 25- to 64-year-olds. The representativeness of driver lists in regard to other factors and in other geographic regions deserves further evaluation. (*Am J Public Health.* 1994;84:469–472)

Introduction

Epidemiology is among many research disciplines that struggle with subject selection for population-based studies. Complete population rosters are rarely available. Without such listings, researchers must identify sampling frames that adequately represent the target population.

In the United States, sampling frames can be compared with the decennial census, stratifying on a few factors that serve as indicators of overall representativeness. We applied this strategy to examine the adequacy of driver's license lists as a sampling frame. We identified age groups with high coverage and then stratified by sex, county, and four levels of county urbanicity to assess representativeness within these groups.

Methods

The Iowa Department of Transportation provided driver's license data in February 1991. We adjusted this data set to reflect resident license holders on April 1, 1990, and then compared it with 1990 census enumerations.[1] Approximately 5% of these records were for nondrivers with picture identification cards.[1]

The original data set contained 2 506 843 records. We eliminated (1) 18 records involving birth dates before January 1, 1870, or after April 1, 1990; (2) 1836 records listing a state of address outside Iowa; (3) 279 records listing nonexistent county codes; (4) 3081 duplicate records; (5) 2313 deceased individuals; (6) 121 887 individuals who had

Charles F. Lynch and Nyla Logsden-Sackett are with the Department of Preventive Medicine and Environmental Health, University of Iowa, Iowa City. Sandra L. Edwards is with the Department of Family and Preventive Medicine, University of Utah, Salt Lake City. Kenneth P. Cantor is with the Environmental Epidemiology Branch, National Cancer Institute, Bethesda, Md.

Requests for reprints should be sent to Charles F. Lynch, MD, PhD, Department of Preventive Mecicine and Environmental Health, 2800 Steindler Bldg, University of Iowa, Iowa City, IA 52242.

This paper was accepted June 29, 1993.

TABLE 1—Populations of Iowans (All Races) by Age Group and Sex: 1990 Driver's License (DL) List vs 1990 Census

	Males			Females		
Age, y	DL List[a]	Census	Ratio[b]	DL List[a]	Census	Ratio[b]
0–4	12	98 754	0.00	8	94 449	0.00
5–9	41	107 301	0.00	38	102 037	0.00
10–14	23 758	104 787	0.23	20 621	99 430	0.21
15–19	99 531	101 398	0.98	94 027	98 018	0.96
20–24	112 781	99 442	1.13	104 910	96 977	1.08
25–29	122 035	102 258	1.19	117 279	104 358	1.12
30–34	120 500	110 293	1.09	116 545	111 710	1.04
35–39	113 574	106 553	1.07	108 491	105 567	1.03
40–44	95 041	91 436	1.04	92 684	91 765	1.01
45–49	73 875	72 161	1.02	72 973	73 891	0.99
50–54	63 278	62 556	1.01	63 813	65 820	0.97
55–59	58 269	58 589	0.99	60 345	63 746	0.95
60–64	59 418	60 405	0.98	61 467	66 948	0.92
65–69	54 174	55 900	0.97	58 576	66 554	0.88
70–74	42 597	44 744	0.95	49 157	59 763	0.82
75–79	28 581	33 255	0.86	34 955	51 231	0.68
80–84	15 535	20 454	0.76	19 420	38 950	0.50
85+	7 207	14 516	0.50	7 797	40 739	0.19
All ages	1 090 207	1 344 802	0.81	1 083 106	1 431 953	0.76

[a]Excludes individuals who died or surrendered their Iowa license for an out-of-state license on or before April 1, 1990, or whose license was expired as of April 1, 1990, and not renewed as of March 1, 1991.
[b]Driver's license count/census count.

surrendered their Iowa licenses by April 1, 1990; and (7) 204 288 individuals whose licenses had expired by April 1, 1990, and were not renewed by March 1, 1991. A total of 333 530 records were dropped from the original data set after meeting one or more of these exclusion criteria, leaving 2 173 313 records.

We stratified the 99 Iowa counties into four mutually exclusive urbanicity groups: (1) 11 counties containing at least one census-defined metropolitan area, (2) 37 counties with no metropolitan areas, but with at least 40% of the population residing in census-defined urban areas, (3) 31 counties with no metropolitan areas and with less than 40% of the population residing in urban areas, and (4) 20 rural counties.

Results

Table 1 compares the adjusted 1990 Iowa driver's license list with the Iowa portion of the 1990 census by 5-year age group and sex. The 0- to 9-year-olds listed have picture identification cards, not driver's licenses. Coverage exceeding 90% is observed among 15- to 74-year-old males and among 15- to 64-year-old females. Overall, coverage decreases gradually with age. In every age group, coverage of males is slightly higher than that of females.

Table 2 shows no substantial differences between county groups except in the 15- to 24-year age range. In this age group, overcounts are observed in the low urbanicity and rural county categories.

Table 3 provides the numbers of individuals who surrendered their Iowa license for a license in another state. Most out-migration occurred among 20- to 29-year-olds.

County-by-county ratios of driver counts to census counts for 25- to 64-year-old men ranged from 0.87 to 1.15. For 25- to 64-year-old women, ratios ranged from 0.90 to 1.08. No regional patterns were evident for either sex.

Discussion

Our results suggest that Iowa driver's license lists adequately represent 25- to 64-year-old residents. Among 15- to 24-year-olds, coverage was high but was related to urbanicity of resident county. Our coverage estimates for 25- to 54-year-olds must be interpreted with caution because we were unable to adjust for driver overestimates (Table 1). Table 3 shows that out-migration occurs most frequently in this same age group, suggesting that out-migration largely accounts for the overestimates. Among 55- to 64-year-olds, out-migration is infrequent, suggesting that coverage estimates are accurate for this age group. We would expect these coverage estimates (above 90%) to be similar among other preretirement age groups (25- to 54-year-olds).

Our conclusions confirm the findings of a 1970s Kentucky study involving stratified sampling on age and sex; in that study, driver lists provided "an excellent" sampling frame for all adult age groups.[2] In addition, the US Department of Transportation compared 1989 Bureau of the Census estimates with 1990 driver's license holders.[3] Of the adult population nationwide, 94.7% of men and 82.8% of women had driver's licenses. Coverage surpassed 90% for men between 20 and 69 years of age and for women between 30 and 49 years old.

Our conclusions also support findings in a recent cancer case–control study[4] in which we selected 40- to 64-year-old controls from Iowa driver's license records. Among 40- to 64-year-old tumor registry patients, 97% of men and 92% of women reported holding driver's licenses. Therefore, selecting controls from driver's records was unlikely to have introduced any substantial selection bias.

The lists were extremely convenient and cost-effective to use. To obtain subject telephone numbers, we used telephone directories, directory assistance, and, for unlisted numbers, a memo to the respondent requesting the telephone number. We successfully contacted 95.8% of the 1077 originally selected driver's license control subjects by these means. We encountered one disadvantage: Iowa license records are updated for address changes only at the 4-year license renewal. This resulted in increased effort in locating some individuals. Limiting the sampling frame to recent license issues was not an effective solution because renewals occur in age cycles, introducing an age bias.

Other population-based sampling frames have provided poor coverage and/or poor representation in some research settings. These include motor vehicle registries,[2] telephone directories,[5–10] county assessors' lists of property owners,[11] commercial city directo-

TABLE 2—Ratios[a] of 1990 Driver's License Holders vs the 1990 Census, by County Urbanicity Group, Sex, and Age Group

	Metropolitan Area		Nonmetropolitan Area				Nonmetropolitan Area	
			High Urbanicity		Low Urbanicity		Rural	
Age, y	Males	Females	Males	Females	Males	Females	Males	Females
0–9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10–14	0.20	0.18	0.24	0.22	0.25	0.22	0.25	0.24
15–19	0.97	0.92	0.95	0.95	1.04	1.05	1.07	1.08
20–24	1.10	1.02	1.04	1.06	1.39	1.33	1.45	1.38
25–29	1.21	1.13	1.17	1.12	1.19	1.11	1.19	1.08
30–34	1.13	1.06	1.07	1.04	1.04	1.01	1.04	0.99
35–39	1.09	1.05	1.06	1.02	1.02	0.99	1.03	0.99
40–44	1.07	1.02	1.02	1.01	1.00	0.98	1.02	0.99
45–49	1.05	1.00	1.01	0.99	0.99	0.97	1.00	0.98
50–54	1.03	0.98	1.00	0.96	0.99	0.96	0.99	0.96
55–59	1.01	0.95	0.99	0.95	0.97	0.93	0.97	0.94
60–64	1.00	0.92	0.97	0.91	0.97	0.92	0.96	0.92
65–69	0.98	0.87	0.96	0.88	0.96	0.89	0.96	0.90
70–74	0.94	0.81	0.95	0.82	0.97	0.83	0.97	0.85
75–79	0.85	0.65	0.86	0.69	0.88	0.70	0.87	0.73
80–84	0.74	0.46	0.76	0.51	0.77	0.51	0.79	0.55
85+	0.50	0.18	0.49	0.19	0.49	0.19	0.51	0.21
All ages	0.82	0.77	0.80	0.75	0.80	0.79	0.81	0.75

Note. The driver's license category excludes individuals who died or surrendered their Iowa license for an out-of-state license on or before April 1, 1990, or whose license had expired as of April 1, 1990, and was not renewed as of March 1991. Metropolitan area counties contain at least one community of 50 000 or more people; nonmetropolitan, high urbanicity counties do not contain metropolitan areas, but more than 40% of the county population lives in urban areas (2500 or more people); nonmetropolitan low urbanicity counties have less than 40% of the county population in urban areas; rural counties contain no metropolitan or urban areas.

[a]Driver's license count/census count.

ries, and voter registries.[2,11] Health Care Financing Administration lists have high coverage nationwide,[12–14] but only for those 65 years of age and older.

Even the frequently used random-digit dialing method[15] underrepresents segments of the population in some locations.[8,16–18] In addition, random-digit dialing involves a large amount of wasted effort calling unassigned and nonresidential numbers.[19–21] Attempts to reduce expense may bias the sample.[22] Furthermore, random-digit dialing usually requires household screening, increasing costs and often reducing response rates.[23]

We believe that driver's license lists are superior to many of these commonly used population-based sampling frames for Iowa. They are available for a nominal fee, computerized, and pooled statewide. New license issues are added promptly. Variables commonly used as eligibility criteria (birth date, sex, address, residence city, and residence county) are contained in license records. Therefore, household screening to select subjects is usually unnecessary, significantly reducing expense and increasing response rates. The vision, literacy, and health requirements for driving may

TABLE 3—Individuals Excluded from the 1990 Driver's License Data Set as a Result of Out-of-State Migration, by Age and Sex

	Out-of-State Migration			
	Male		Female	
Age y	No.	%[a]	No.	%
0–9	0	0.0	0	0.0
10–14	1	0.0	0	0.0
15–19	2 348	2.2	2 120	2.2
20–24	11 198	8.4	12 202	10.8
25–29	15 563	9.9	14 179	9.6
30–34	9 785	6.2	8 880	6.3
35–39	7 306	5.4	6 751	5.4
40–44	5 323	4.8	4 903	4.7
45–49	3 442	4.1	3 062	3.8
50–54	2 151	3.1	1 954	2.8
55–59	1 678	2.7	1 613	2.5
60–64	1 507	2.4	1 295	2.0
65–69	1 230	2.1	1 051	1.6
70–74	646	1.1	595	1.1
75–79	347	1.1	343	0.9
80–84	143	0.8	174	0.7
85+	60	0.5	37	0.2
Total	62 728	5.0	59 159	4.8

Note. Out-of-state migrators surrendered their Iowa licenses for out-of-state licenses on or before April 1, 1990.

[a]Percentage of individuals listed on the original driver's license list but subsequently excluded for out-of-state migration.

exclude some individuals from the lists. However, as mentioned, the lists also contain nondrivers who obtain picture-identification cards.

Our conclusions are based solely on comparisons of Iowa drivers with Iowa census enumerations. The accessibility and representativeness of driver's license records in other regions may differ. □

Acknowledgment

This research was supported in part by a Public Health Service Preventive Oncology Academic Award (5 KO7 CA01181-05).

References

1. *1990 Census of Population and Housing: Summary Tape File 1A: State of Iowa.* Washington, DC: US Bureau of the Census; 1990.
2. Warner PD, Hoffman SD, Burdge RJ. Driver's license lists as a sampling frame. *Public Data Use.* 1977;5:3–10.
3. *Driver Licenses—1990.* Washington, DC: US Dept of Transportation; 1990. FHWA-PL-92-002.
4. Cantor KP, Lynch CF, Johnson D. Bladder cancer, parity, and age at first birth. *Cancer Cause Control.* 1992;3:57–62.
5. Brunner JA, Brunner GA. Are voluntarily unlisted telephone subscribers really different? *J Marketing Res.* 1971;8:121–124.
6. Leuthold DA, Scheele R. Patterns of bias in samples based on telephone directories. *Public Opinion Q.* 1971;35:249–257.
7. Moberg PE. Biases in unlisted phone numbers. *J Advertising Res.* 1982;22:51–55.
8. Corey CR, Freeman HE. Use of telephone interviewing in health care research. *Health Serv Res.* 1990;25:129–144.
9. Orden SR, Dyer AR, Liu K, et al. Random digit dialing in Chicago CARDIA: comparison of individuals with unlisted and listed telephone numbers. *Am J Epidemiol.* 1992;135:697–709.
10. Marcus AC, Crane LA. Telephone surveys in public health research. *Med Care.* 1986;24:97–112.
11. Stanek EJ, Pastides H, Darity WA, Elkins M. City directories as sources for survey work in low- and middle-income Black communities. *Am J Public Health.* 1990;80:1125–1126.
12. Hartge P, Cahill JI, West D, et al. Design and methods in a multi-center case-control interview study. *Am J Public Health.* 1984;74:52–56.
13. Hatten J. Medicare's common denominator: the covered population. *Health Care Financing Rev.* 1980;2(2):53–64.
14. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies: II. Types of controls. *Am J Epidemiol.* 1992;135:1029–1041.
15. Waksberg J. Sampling methods for random digit dialing. *J Am Stat Assoc.* 1978;73:40–46.
16. Davis S, Ross A, Voigt LF, Heuser L. Qualitative and quantitative assessment of geographic clustering of population samples selected using different methods of random digit dialing. *Am J Epidemiol.* 1990;132:S144–S155.
17. Thornberry OT, Massey JT. Coverage and response in random digit dialed national surveys. In: *Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association; 1983:654–659.
18. Thornberry OT, Massey JT. Trends in United States telephone coverage across time and subgroups. In: Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WL, Waksberg J, eds. *Telephone Survey Methodology.* New York, NY: Wiley; 1988.
19. Franz JD. Response rates in random digit dialed telephone surveys: estimation vs. measurement. Presented at the annual meeting of the 67th American Educational Research Association; April 16–20, 1986; San Francisco, Calif.
20. Dillman DA. Implementing telephone surveys. In: *Mail and Telephone Surveys: The Total Design Method.* New York, NY: Wiley; 1978.
21. Groves RH, Kahn RL. *Surveys by Telephone.* New York, NY: Academic Press; 1979.
22. Voigt LF, Davis S, Heuser L. Random digit dialing: the potential effect on sample characteristics of the conversion of nonresidential telephone numbers. *Am J Epidemiol.* 1992;136:1393–1399.
23. Psaty BM, Cheadle A, Curry S, et al. Sampling elderly in the community: a comparison of commercial telemarketing lists and random digit dialing techniques for assessing health behaviors and health status. *Am J Epidemiol.* 1991;134:96–106.

Neural Tube Defects in Jamaica following Hurricane Gilbert

E. Marilyn Watson Duff, MScNutr, BScN, and Edward S. Cooper, MB, BS, FRCP

ABSTRACT

An increased incidence of obvious live-birth neural tube defects (i.e., spina bifida cystica and encephalocele) occurred in Jamaica 11 to 18 months after Hurricane Gilbert. The conceptions of the affected babies coincided with a rise in megaloblastic change in sickle cell patients, suggesting a widespread drop in dietary folate intake. A detailed history was taken from each of the 17 affected mothers (case subjects) and 51 unaffected mothers (matched control subjects). The case subjects reported a significantly lower mean intake of dietary folate in the periconceptional period (154 μg/day) than did the control subjects (254 μg/day). The temporary increase in neural tube defects was associated with a diet comparatively low in folate in the periconceptional period, suggesting the dietary level of folate that fails to protect against neural tube defects under natural conditions. (*Am J Public Health.* 1994;84: 473–476)

Introduction

In 1991 the Medical Research Council Vitamin Study showed that recurrence of neural tube defects could be prevented by daily supplementation of the diets of high-risk mothers with 4 mg of folate around the time of conception.[1]

After the publication of this report, there was controversy over the dosage of folate supplementation and particularly the possible toxic effects of circulating unmetabolized free folate on the developing embryo.[2–6] In addition, differences of opinion existed worldwide as to what may be considered a folate-replete diet.[6–9]

In Jamaica, the incidence of obvious live-birth neural tube defects (spina bifida cystica and encephalocele) is low, averaging 1.4 per 10 000 live births over 10 years (an average of 2.0 referrals per quarter). In the aftermath of Hurricane Gilbert (September 12, 1988) there was an increase in live-birth neural tube defects with a peak of 5.0 per 10 000 (7 referrals) in one quarter. This increase, which has been reported elsewhere,[10] was significant (Poisson model; $P < .005$) and coincided in periconceptional time with an overt, significant increase in folic acid deficiency in a vulnerable group: homozygous sickle cell patients.[11]

The overwhelming, all-island devastation in the aftermath of Hurricane Gilbert greatly enhanced recall of events around that time. The results presented here, despite the retrospective nature of the information and the small numbers, shed light on the dosage controversy and the range of daily dietary folate that is inadequate to prevent neural tube defects.

Methods

The cases were the 17 mothers of babies with obvious live-birth neural tube defects (spina bifida cystica and encephalocele) born in the third and fourth quarters of 1989 and the first quarter of 1990 who were referred for treatment within the Jamaican hospital system. All cases of neural tube defects were clinically confirmed in tertiary institutions. Anencephaly was excluded because such cases are not usually referred and reports of this diagnosis could not be confirmed. No record of an anencephalic live birth or stillbirth was available from the main hospitals nor from the Ministry of Health. The recording of such birth defects is known to be deficient.[12]

Each case subject was matched with three control subjects, mothers of babies with a birthweight of >2.5 kg and without any apparent physical or developmental defects. Case and control subjects were matched for maternal age, month and year of conception, periconceptional geographic location (Figure 1), parity, and educational level. All of the subjects were of similar socioeconomic and ethnic background (African descent).

A 56-item questionnaire was used in a structured personal interview of each of the case and control subjects by the sole investigator (E.M.W.D.). Blinding of the identities of the case and control subjects was not feasible. Each interview was completed within 11–18 months of the subject baby's birth. The questionnaire was developed, pretested, and precoded to explore 32 health, environmental, and dietary variables in the postdisaster periconceptional period of each subject's pregnancy (see Appendix). Question 55 consisted of a "food folate content scale" that listed 34 common food items and their estimated folate content, designed to give an approximate indication of the subject's daily intake of free folate (in micrograms) in that period for comparison between case and control subjects. This scale was developed from food composition tables[13] and prorated for local dietary servings.

The authors are with the Tropical Metabolism Research Unit, University of the West Indies, Jamaica, West Indies.

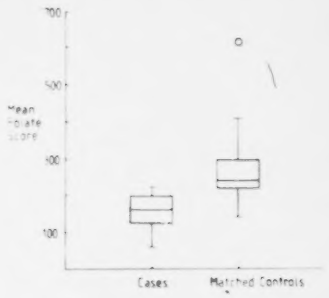
Requests for reprints should be sent to E. Marilyn Watson Duff, MScNutr, BScN, Tropical Metabolism Research Unit, University of the West Indies, Mona, Kingston 7, Jamaica, West Indies.

This paper was accepted September 7, 1993.

Editor's Note. See related annotation by Beresford (p 348) in this issue.

Note: Each circle has a diameter of approximately 8 kilometers.

FIGURE 1—The periconceptional geographic locations in Jamaica of the case subjects and their matched control subjects.



Note. The high outlying score was that of the only subject (control) who took a daily supplement.

FIGURE 2—The mean folate intake scores (micrograms/day) for the two groups, the upper and lower limits, and the upper and lower quartiles of the two distributions.

TABLE 1—Mean Daily Folate Intake Reported in the Periconceptional (Post-Hurricane) Periods of Neural Tube Defect Case Subjects and Control Subjects

Group	No.	Mean ± SD Daily Folate Intake (μg)	95% CI
Case subjects	17	153.6 ± 49.3	123, 185
Control subjects	50[a]	253.9 ± 64.3	233, 275

Note. $t = -5.86$; $df = 65$; $P < .0001$.
[a]One high outlying score was omitted from analysis.

The periconceptional period was defined as an 18-week period starting 12 weeks before conception and ending 6 weeks after conception.

Information on postdisaster food availability and relief supplies was obtained from the Salvation Army, the main relief agency.

The mean folate scores of group 1 (case subjects) and group 2 (control subjects) were compared by using the *t* test for independent samples. Fisher's Exact Test and the corrected chi-square test were also used where applicable. Possible confounding variables such as minor trauma and oral contraceptive use were controlled for with multiple logistic regression analysis.

Results

The mean daily dietary intake of folate was significantly lower ($P < .0001$) in group 1 (case subjects) (154 μg/day, 95% confidence interval [CI] = 123, 185) than in group 2 (control subjects) (254 μg/day, 95% CI = 233, 275) (Table 1, Figure 2). The range for group 1 was 60–220 μg/day. The range for group 2 was 140–410 μg/day with a high outlying intake of 616 μg/day, that of the only mother (control subject) who in addition to a dietary intake of 445 μg/day supplemented herself with a multivitamin "liquid food" preparation containing folic acid (approximately 170 μg/day). For each case subject the mean folate intake of the matched control subjects was higher (Figure 3).

There was no association with maternal smoking; alcohol, marijuana, or other substance abuse; pesticide exposure; intake of previously suspected teratogenic foods (i.e., ackee,[14] potato,[15] cassava, Jamaican white yam[16]); bush teas; pica; sources of drinking water; clinic attendance; environmental or psychosocial stress; hyperemesis; hypertension; spacing of pregnancies; incestuous parenting; or previous birth defects.

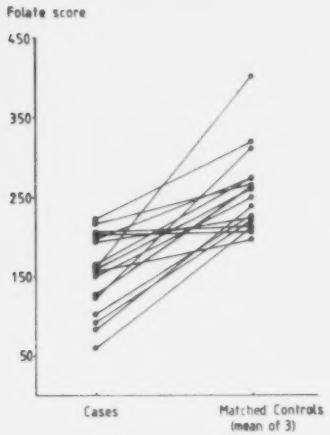
There was no apparent relationship with maternal medication use, but oral contraceptive use[17] was higher among the case subjects, and this difference approached statistical significance (odds ratio [OR] = 3.8, 95% CI = 0.95, 15.44, $P = .058$).

No previous birth defects were reported by any of the subjects in either group. Stillbirths were reported to have occurred in previous pregnancies of two of the case subjects and none of the control subjects. Trauma (minor trauma from blows and falls) was reported only by subjects in group 1 (OR = 5.3, 95% CI = 3.16, 8.73, $P < .0005$). There was no apparent explanation for this.

Population fertility (live births) increased by 4.3% over the same months of the previous year (Statistical Institute of Jamaica, 1988–1990, unpublished data). Although there was no general shortage of food reported by the subjects, there was a change in eating patterns at that time, towards the use of prepared and packaged foods (relief supplies consisting of rice, flour, sugar, cornmeal, tinned corned beef, sausages, sardines, and mackerel) with a considerably reduced intake of fresh produce.

Discussion

We recognize the limitations of retrospective data and a study in which blinding of case and control identities was not practicable, but the special opportunity presented by this experiment of nature mandated the investigation.



Note. In each instance the mean folate intake of the three matched control subjects was higher than that of the case subject.

FIGURE 3—The dietary folate intake scores (micrograms/day) of each case subject compared with the mean dietary folate intake scores of the three matched control subjects.

Within these limitations, the results show that this post-Hurricane Gilbert increase in neural tube defects was associated with a diet comparatively low in folate in the periconceptional period.

The range of reported daily dietary folate intakes among the case subjects (60–220 μg/day) and among the control subjects (140–410 μg/day) gives an approximate indication of the folate content (i.e., micrograms of free folate daily) of the individual diets of the mothers in their post-hurricane periconceptional periods. From this result it may be suggested that a dietary intake of <220 μg of folate per day in the periconceptional period may be inadequate to prevent neural tube defects in the babies of Jamaican mothers. No estimate of the converse, the intake that is protective, can come from a case–control study such as this one.

The MRC Vitamin Study Research Group suggested that the United Kingdom may have a lower folic acid intake than other countries.[1] The usual daily mean intake of folate for British women is 209 μg/day.[8]

The U.S. recommended daily allowance for folate is 180 μg/day for nonpregnant women and 400 μg/day during pregnancy. In the 1976–1980 US Health and Nutrition Survey, median folate intakes were 166 μg/day and 146 μg/day for White and Black women, respectively.[6]

Scott et al.[8] and Leeming[4] emphasized the danger of exposure of the developing embryo to circulating unmetabolized folic acid arising from high-dose supplementation. They recommend that a woman contemplating pregnancy could avoid neural tube defects in her baby by either taking a low-level folic acid preparation or eating an improved or even fortified diet high in folic acid. Strieff[18] suggested that orange juice was a good dietary supplement for folic acid. This suggestion was endorsed by Lucock et al.,[3] who demonstrated in their experiments an enhanced absorption and metabolism of folic acid after ingestion of 500 mL of pure orange juice.

In view of possible adverse reactions to high-dose folic acid supplementation and the recognized interrelation of nutrients in a varied diet adequate in folic acid and related nutrients (such as ascorbate, pyridoxine, and cobalamin), it has been suggested that dietary sources of folate may be preferable to high-dose supplementation.[7]

The Jamaican diet usually includes a wide variety of tropical fruits and vegetables throughout the year. Many of these (e.g., ripe bananas and oranges) contain appreciable amounts of folate and ascorbate. Chicken and beef liver and kidney, eggs, callaloo (a local vegetable similar to spinach), avocados, and steamed rice are also popular in the Jamaican diet. That may account for the usually low incidence of neural tube defects in the Jamaican population.

The massive destruction of crops and livestock wrought by Hurricane Gilbert had immediate impact on the diets of the people of Jamaica. Fresh produce was scarce and therefore expensive.[11] Many of the mothers reported returning to their usual diets 3 to 4 months after the hurricane. □

Acknowledgments

We thank the Commonwealth Caribbean Medical Research Council for partial funding of this research. Edward S. Cooper is a Wellcome Trust Senior Clinical Research Fellow.

The abstract of this paper was presented at the 37th Scientific Meeting of the Commonwealth Caribbean Medical Research Council, April 22–25, 1992, in Curacao, Netherlands Antilles.

We also thank Dr Donald Simeon of the Tropical Metabolism Research Unit for assistance with the statistical analyses.

References

1. MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *Lancet.* 1991;338:131–136.
2. Smithells RW, Sheppard S, Schorah CJ, Wild J. Folic acid to prevent NTDs. *Lancet.* 1991;338:379–380. Letter.
3. Lucock MD, Wild J, Hartley R, Levene MI, Schorah CJ. Vitamins to prevent neural tube defects. *Lancet.* 1991;338:894–895. Letter.
4. Leeming RJ. Vitamins to prevent neural tube defects. *Lancet.* 1991;338:895. Letter.
5. Reynolds EH. Folic acid to prevent neural tube defects. *Lancet.* 1991;338:506. Letter.
6. Folate supplements prevent recurrence of neural tube defects. *Nutr Rev.* 1992;50:22–24.
7. Crawford MA. Folic acid to prevent neural tube defects. *Lancet.* 1991;338:379–380. Letter.
8. Scott JM, Kirke P, O'Broin S, Weir DG. Folic acid to prevent neural tube defects. *Lancet.* 1991;338:505. Letter.
9. Willett WC. Folic acid and neural tube defects: can't we come to a closure? *Am J Public Health.* 1992;82:666–668.
10. Duff EMW, Cooper ES, Danbury CM, Johnson BE, Serjeant GR. Neural tube defects in hurricane aftermath. *Lancet.* 1991;337:120–121. Letter.
11. Readett DRJ, Serjeant BE, Serjeant GR. Hurricane Gilbert anaemia. *Lancet.* 1989;2:101–102. Letter.
12. Coard K, Escoffery C, Golding J, Ashley D. Incidence of anencephaly in Jamaica. *Teratology.* 1990;41:173–176.
13. Paul AA, Southgate DAT. *McCance & Widdowson's The Composition of Foods.* London, England: Her Majesty's Stationery Office; 1978.
14. Persaud TVN. Foetal abnormalities caused by the active principle of the Blighia sapida (ackee). *West Indian Med J.* 1967;16:193–197.
15. Renwick JH. Spina bifida, anencephalus and potato blight. *Lancet.* 1972;2:967–968.
16. Golding J, Foster-Williams K, Coard K, Ashley DA. Cluster of central nervous system defects in Jamaica. *Hum Exp Toxicol.* 1990;9:13–16.
17. Davis RE. Folic acid to prevent neural tube defects. *Lancet.* 1991;338:506. Letter.
18. Strieff RR. Folate levels in citrus and other juices. *Am J Clin Nutr.* 1971;24:1390.

(*Appendix on next page*)

APPENDIX—List of Variables Investigated by the Structured Personal Interview

1. *Personal information.* Subject's name, address, age, and educational level; occupations of subject and spouse; socioeconomic status (by employment skills rating); family structure; smoking; medication; substance use and abuse.
2. *Antenatal care.* Parity, previous stillbirths and/or birth defects, contraceptive use, spacing, general health, illnesses, accidents, results of diagnostic tests, medication, supplements, clinic attendance.
3. *Index baby.* Name; sex; date of birth; gestation; birth weight; month, year, and place of conception; place and type of delivery; health status at birth; admission(s) to hospital, diagnosis, treatment, and outcome.
4. *Hurricane experiences.* Subject's personal experiences in the aftermath of Hurricane Gilbert. Physical and psychosocial stress, damage, relocation, available food, drinking water.
5. *Periconceptional period.* Nutritional status (as described by the subject), supplements, medications, oral contraceptive use, illnesses, accidents, smoking, substance use and abuse, use of "bush teas," exposure to pesticides, consumption of previously suspected teratogenic foods (i.e., potato, Jamaican white yam, ackee, and/or cassava), pica, unusual stressful events. Exploration of periconceptional diet, determination of the "folate score."

The Use of Commercial Tanning Facilities by Suburban Minnesota Adolescents

Jennifer A. Oliphant, MPH, Jean L. Forster, PhD, MPH, and Colleen M. McBride, PhD

ABSTRACT

In 1991, 1008 suburban St. Paul, Minn, high school students were surveyed via self-administered questionnaire regarding use of commercial tanning facilities, injuries experienced from tanning, use of protective measures while tanning indoors, and knowledge of the risks of tanning. Overall, 34% of the respondents had used commercial tanning facilities. Fifty percent said they had not been warned by tanning facility operators about the risks of tanning indoors, 28% reported not being told to wear goggles, and 17% reported never wearing goggles. The results indicate that these adolescents use commercial tanning services at high rates, and often in ways that increase their risk for a variety of health problems. (*Am J Public Health*. 1994;84:476–478)

Introduction

Exposure to ultraviolet radiation, such as that from artificial tanning devices, is a risk factor for a variety of health problems including erythema, vesiculation, photosensitizing reactions, actinic elastosis, keratoses, basal cell and/or squamous cell carcinoma, and malignant melanoma.[1–3] Exposure to ultraviolet radiation can also result in compromised immune response, eye burns, and cataracts.

The risk of many health problems is higher when ultraviolet radiation exposure occurs prior to adulthood.[4] The damage begins with a child's first exposure and accumulates through his or her lifetime.[5,6] It is estimated that up to 78% of the risk from ultraviolet radiation exposure is completed by age 18.[7]

Indoor tanning is not safer than tanning in natural sunlight. Commercial tanning devices emit either mostly ultraviolet A or mostly ultraviolet B.[2,8,9] Although ultraviolet B injures the skin faster, ultraviolet A penetrates the skin more deeply.[9,10] Some commercial tanning beds may emit as much as 10 times more ultraviolet A than natural sunlight.[2]

Artificial tanning produces no known health benefits,[3] nor is it thought possible to tan safely.[6] Exposure to ultraviolet radiation from indoor tanning increases the chances of developing cancer from natural sunlight.[11,12] The sun protection factor from tanning indoors is only about 4, which offers no significant protection from sun exposure.[13] Moreover, severe ultraviolet radiation burns (those that cause pain for 48 hours or more[14]) increase the risk for malignant melanoma.[4,15]

Despite the serious risks associated with artificial tanning, fewer than half of the states currently regulate commercial tanning facilities. The US Food and Drug Administration (FDA) regulation requires only that the device be equipped with a calibrated timer, protective eyewear, and a small warning label.[16] A previous study found that 45% of all commercial tanning devices examined

Jennifer A. Oliphant is with the Johnson County Department of Public Health, Iowa City, Iowa. Jean L. Forster is with the University of Minnesota School of Public Health, Minneapolis. Colleen M. McBride is with the Center for Health Studies, Group Health Cooperative, Seattle, Wash.

Requests for reprints should be sent to Jean L. Forster, PhD, MPH, University of Minnesota School of Public Health, Division of Epidemiology, Suite 300, 1300 S 2nd St, Minneapolis, MN 55454-1015.

This paper was accepted June 3, 1993.

TABLE 1—Lifetime Prevalence of Use of Commercial Tanning Facilities

	n[a]	Lifetime Prevalence, %
Gender		
Male	469	15
Female	511	51
Age		
14 y	135	18
15 y	303	29
16 y	261	34
≥ 17 y	283	46
Total	987	34

[a]These n's reflect missing values for gender (28), age (26), or tanning experience (21).

TABLE 2—Reported Age of First Use of Commercial Tanning Facility, by Age of User

		Age of First Use, %			
Age of User	n[a]	9–11 y	12–14 y	15–17 y	> 17 y
14 y	22	. . .	100	. . .	. . .
15 y	80	8	81	11	. . .
16 y	88	. . .	53	47	. . .
≥ 17 y	111	2	22	74	2

[a]These n's reflect the fact that 30 of the 331 tanners had missing values for either age of first use or age of respondent.

did not have the mandated FDA warning label and 50% of the commercial facilities surveyed had at least one improperly labeled machine.[17]

Despite the risks of artificial tanning, little information is available concerning use of tanning devices among the general population and specifically among adolescents. The purpose of this study was to characterize for the first time the prevalence and patterns of use of indoor tanning facilities by adolescents, the health problems associated with adolescents' use of commercial tanning equipment, and adolescents' knowledge of the risks associated with indoor tanning.

Methods

The study sample consisted of all students in grades 9 through 12 from a suburban high school in the St. Paul, Minn, metropolitan area. The survey was self-administered during homeroom and took approximately 5 minutes to complete. A copy of the instrument is available from the author on request. The response rate was 84% (13% absent or missing, 2% refusing, and 1% insincere reporting), yielding a final n of 1008. The sample resembled the Twin Cities suburban population; 88% of the respondents were White. Females slightly outnumbered males (51% to 48%). Participants' ages ranged from 13 to 19 years.

In the survey, students were asked questions to determine their skin type, their history of sun exposure, and their current and past use of commercial tanning equipment. In addition, they were asked to report problems experienced as a result of tanning indoors, their use of protective eyewear, and information received from tanning facility operators regarding the risks of indoor tanning. Questions to ascertain the students' knowledge about the risks of tanning indoors were also asked.

Chi-square tests for independence were used to determine whether high-risk tanning behaviors were associated with each other.

Results

Overall, 34% of the students reported using a tanning facility (Table 1). The lifetime prevalence of indoor tanning was 51% for females and 15% for males. Almost 20% of those aged 14 years or younger reported using a tanning facility; those aged 15, 16, and 17 years or older reported 29%, 34%, and 46% use, respectively. Students' age of first indoor tanning ranged from 9 to 18 years, with more than half beginning before age 15 (Table 2). Those who reported at least one exposure to indoor tanning (n = 331) will be referred to as "tanners."

Most tanners (72%) reported indoor tanning infrequently (less than once a month or on special occasions). However, 28% of all tanners tanned at least once per month and 15% reported tanning once a week or more (Table 3).

The reported usual length of a tanning session ranged from 5 minutes to more than 60 minutes. The majority of tanners reported sessions lasting from 21 to 30 minutes (Table 3). Most reported their longest session as 30 minutes. However, 11% of the tanners reported tanning indoors for more than 30 minutes. More than 40% of the tanners reported not always wearing goggles and 17% reported never wearing them. Chi-square tests of association revealed that those who reported longer usual tanning sessions were less likely to use goggles ($P = .06$) and more likely to tan frequently ($P = .02$).

TABLE 3—Practices of Users of Commercial Tanning Facilities

	%
Average length of tanning session (n = 322)	
1–10 min	4
11–20 min	30
21–30 min	63
> 30 min	3
Longest single session (n = 325)	
< 30 min	43
30 min	46
> 30 min	11
Use of goggles while tanning (n = 327)	
Always	59
Usually	11
Sometimes	13
Never	17
Frequency of sessions (n = 326)	
≥ 1/week	15
< 1/week but ≥ 1/month	13
< 1/month	72

Skin and eye problems were frequently reported by indoor tanners. Approximately 16% of the tanners had experienced one or more eye problems, including pinkeye, eye burns, or sandy or gritty eyes. Fifty-nine percent reported some skin injury, including burned, blistered, or peeled skin and/or rashes.

Tanners were asked to report the instructions and warnings they received

TABLE 4—Reported Instructions and Warnings from Tanning Facility Attendant

	%
Told to wear goggles (n = 314)	
Always	42
Usually	16
Sometimes	14
Never	28
Told to limit length of session (n = 314)	
Always	35
Usually	13
Sometimes	25
Never	26
Received warning about the health risks of indoor tanning (n = 305)	
Yes, written	30
Yes, verbal	9
Yes, both	11
Never	50
Ever noticed a sign warning of the risks of indoor tanning (n = 301)	48

from operators at the business where they tanned most often. Table 4 shows that more than half were not always told to wear goggles, more than a fourth said they were never told to limit their time per session, half reported that they had never received a warning about the health risks of indoor tanning, and fewer than half had ever noticed a sign warning of the health risks of tanning.

All students were asked to respond to true/false statements regarding the risks of indoor tanning. Sixty-one percent knew that closing their eyes or covering them with cotton balls while tanning was not enough protection, 71% knew that indoor tanning is not safer than natural sunlight, 79% knew that indoor tanning could cause skin cancer, and 77% were aware that damage could occur from tanning even without a sunburn. Only 22% of the students knew that tanning is more harmful to adolescents than adults. The tanners were as knowledgeable as those who had never tanned indoors.

Discussion

Our results indicate that these adolescents, especially females, use commercial tanning devices at high rates, begin tanning indoors at young ages, and often do not engage in protective behaviors. Especially alarming is the number of weekly tanners (15% of tanners) and the number who infrequently or never use goggles while tanning indoors (30% of tanners). Furthermore, it appears that tanning facility operators often do not warn users of the risks of tanning, and the FDA warning is not noticeable to the majority of adolescent tanners.

In Minnesota communities, commercial tanning facilities are readily available. For instance, 33 tanning businesses with a total of 150 tanning beds were located in the suburban area where students included in this survey live, or 1 tanning bed for every 1200 residents.

Despite the health risks associated with indoor tanning during adolescence, many commercial tanning facilities in Minnesota target this population. Indoor tanning facilities often offer a variety of student discounts and promotions, such as price reductions, multiple tanning session packages, advertisements in school yearbooks and student newspapers, and free tanning coupons as prizes at school functions.

Of the 22 states that regulate commercial tanning facilities, only about half require parental consent for use of commercial tanning facilities, and none prohibit their use by minors. This is despite the recommendations made by the American Academy of Dermatology,[3] the American Medical Association,[5] the Skin Cancer Foundation,[18] the American Cancer Society,[19] and the United States Public Interest Research Group[17] for prohibition or severe limitations of use by minors and adults.

The limitations of this study include the use of self-reports about tanning activity, the fact that reports of medical problems were not validated, and the use of a single suburban school as the only data collection site, which limits generalizability. However, these data indicate the potential for widespread exposure to serious health risks in this population. Further studies are needed to explore the prevalence of adolescent use nationwide and to document acute and chronic injury caused by the use of commercial tanning devices. □

Acknowledgments

This project was supported by the Berglund Foundation and by internal funds made available by the University of Minnesota.

References

1. *Indoor Tanning: Facts for Consumers.* Washington, DC: Federal Trade Commission, Bureau of Consumer Protection; September 1988.
2. *The Darker Side of Indoor Tanning: Skin Cancer, Eye Damage, Skin Aging, Allergic Reactions.* Rockville, Md: US Public Health Service; 1987. DHHS publication FDA 87-8270.
3. Bickers DR, Epstein JH, Fitzpatrick TB, Harber LC, Pathak M, Urbach F. Risks and benefits from high-intensity ultraviolet A sources used for cosmetic purposes (special report). *J Am Acad Dermatol.* 1985;12:380–381.
4. Weinstock MA, Colditz GA, Willett WC, et al. Nonfamilial cutaneous melanoma incidence in women associated with sun exposure before 20 years of age. *Pediatrics.* 1989;84:199–204.
5. Council on Scientific Affairs. Harmful effects of ultraviolet radiation. *JAMA.* 1989;262:380–384.
6. Parrish JA. Ultraviolet radiation for skin tanning. *JAMA.* 1981;245:1470. Letter.
7. Stern RS, Weinstein MC, Baker SG. Risk reduction for nonmelanoma skin cancer with childhood sunscreen use. *Arch Dermatol.* 1986;122:537–545.
8. Bryneel-Rapp F, Dorsey SB, Guin JD. The tanning salon: an area survey of equipment, procedures, and practices. *J Am Acad Dermatol.* 1988;18:1030–1038.
9. Hurwitz S. The sun and sunscreen protection: recommendations for children. *J Dermatol Surg Oncol.* 1988;14:657–660.
10. National Cancer Institute. *Nonmelanoma Skin Cancers, Basal and Squamous Cell Carcinomas: Research Report.* Bethesda, Md: US Dept of Health and Human Services; 1988. NIH publication 88-2977.
11. Staberg B, Wulf HC, Klemp P, Poulsen T, Brodthagen H. The carcinogenic effect of UVA irradiation. *J Invest Dermatol.* 1983; 81:524–527.
12. Willis I, Menter JM, Whyte HJ. The rapid induction of cancers in the hairless mouse utilizing the principle of photoaugmentation. *J Invest Dermatol.* 1981;76:404–408.
13. Devgun MS, Johnson BE, Paterson CR. Tanning, protection against sunburn and vitamin D formation with a UV-A 'sunbed.' *Br J Dermatol.* 1982;107:275–284.
14. Green A, Siskind V, Bain C, Alexander C. Sunburn and malignant melanoma. *Br J Cancer.* 1985;51:393–397.
15. Lew RA, Sober AJ, Cook N, Marvell R, Fitzpatrick TB. Sun exposure habits in patients with cutaneous melanoma: a case control study. *J Dermatol Surg Oncol.* 1983;9:981–986.
16. Department of Health and Human Services, Food and Drug Administration. 1040, Sunlamp products: performance standard: final rule (21 CFR 1040). *Federal Register.* 1985;50:36548–36552.
17. Sikes L. *Indecent Exposure: Consumers Getting "Burned" by Tanning Salons.* Washington, DC: US Public Interest Research Groups; 1991.
18. *Simple Guidelines to Help Protect You from the Damaging Rays of the Sun.* New York, NY: The Skin Cancer Foundation; 1992.
19. *Fry Now. Pay Later.* Atlanta, Ga: American Cancer Society; 1986. ACS publication 2611-LE.

Screening for Insecticide Overexposure under Field Conditions: A Reevaluation of the Tintometric Cholinesterase Kit

Rob McConnell, MD, and Ralph Magnotti, PhD

ABSTRACT

A semiquantitative tintometric field kit has been used in the developing world for almost 30 years to measure whole blood cholinesterase levels in persons exposed to organophosphate pesticides. The validity of this screening kit was evaluated among 79 workers heavily exposed to organophosphates by comparison with a reference assay for erythrocyte cholinesterase. Overall correlation between the two methods was good. However, either sensitivity or specificity of the tintometric kit was less than 75% for each of the three tintometric categories commonly used to define the limit of normal. Because baseline erythrocyte cholinesterase levels were not available for this population, the true sensitivity and specificity of the tintometric assay may be even lower. (*Am J Public Health*. 1994;84:479–481)

Introduction

Worldwide, it is estimated that there are approximately 3 million severe poisonings and 220 000 deaths every year due to pesticides.[1,2] Ninety percent of these poisonings and 99% of the deaths occur in developing countries, although these countries account for only 25% of world pesticide consumption.[2]

The measurement of cholinesterase in blood was one of the first biomarkers available as a screening test for a toxic exposure. It is used to identify asymptomatic workers who are overexposed to acutely toxic organophosphate insecticides.[3] In developing countries, an economical, "tintometric" field kit has been available for more than 40 years to measure whole blood cholinesterase activity semiquantitatively in nine intervals (from 0 to 100) as a percentage of the activity of blood from an unexposed control.[4] A previous evaluation of this kit concluded that differences from a standard laboratory assay were "not beyond those which might be anticipated as due to inherent differences in methods."[5] However, the sensitivity and specificity of the kit and its accuracy in identifying workers with mildly depressed erythrocyte cholinesterase were not presented.

In this study we reevaluate the tintometric kit, using as a reference assay a modification of the standard laboratory method (Ellman et al.'s[6]), as adapted for field evaluation of erythrocyte cholinesterase. The diagnostic validity of this assay has been established by laboratory tests and field trials with normal and pesticide-exposed populations.[7,8] In these studies, the assay exhibited a linear response as a function of erythrocyte cholinesterase concentration from 10% to more than 200% of mean normal activity. These measurements were proportional to the response observed on identical codeterminate samples using the method of Ellman et al.[6] ($R > .99$), which is considered to be the standard for cholinesterase assay.[9]

Methods

Seventy-nine male Hispanic workers at rural landing strips and at one large crop-dusting airport in Northern Pacific Nicaragua were examined. These workers' exposures to a wide variety of organophosphate insecticides have been reported elsewhere.[10,11] Each worker was instructed to wash his hands. Then the index finger or thumb was washed with 0.1 normal sodium hydroxide, wiped with 70% ethanol, and allowed to dry. A deep puncture was made with a lancet, and blood was drawn and assayed for whole blood cholinesterase (using the tintometric kit) and for erythrocyte cholinesterase (using the reference assay).

The protocol specified by the manufacturer of the tintometric kit (Tintometer Company, Williamsburg, Va) was followed:[12] the cholinesterase activity of each subject was compared with that of one of two adult male controls, who were unexposed to pesticides. The manufacturer makes the following recommendations for interpretation of the results:

- 100% to 75% of control cholinesterase activity: no action necessary.
- 75% to 50% of control cholinesterase activity: overexposure probable; repeat test. If results are confirmed, suspend subject from further work with organophos-

Rob McConnell is with the Division of Occupational and Environmental Medicine at Mount Sinai School of Medicine, New York, NY. Ralph Magnotti is with the Department of Pathology and Laboratory Medicine, the University of Cincinnati Medical School, Ohio.

Requests for reprints should be sent to Rob McConnell, MD, Division of Occupational and Environmental Medicine, Box 1057, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029.

This paper was accepted September 30, 1993.

TABLE 1—Mean Colorimetric Erythrocyte Cholinesterase,[a] by Tintometric Category: Crop-Dusting Airport Workers

Tintometric Result (Percentage of Control Activity)	No.	Low by Reference Assay		Mean Erythrocyte Cholinesterase (SD)	
		No.	%		
0.0	1	1	100	0.73	(0)
12.5	5	5	100	0.83	(0.28)
25.0	8	8	100	1.5	(0.6)
37.5	4	4	100	2.3	(1.0)
50.0	5	4	80	3.1	(0.4)
62.5	9	6	67	3.4	(0.7)
75.0	18	7	39	3.8	(0.6)
87.5	10	2	20	4.0	(0.4)
100.0	19	1	5	4.6	(0.5)
Within groups total	79			3.4	(0.6)

Note. Spearman R = .85.
[a]Reference assay in international units.

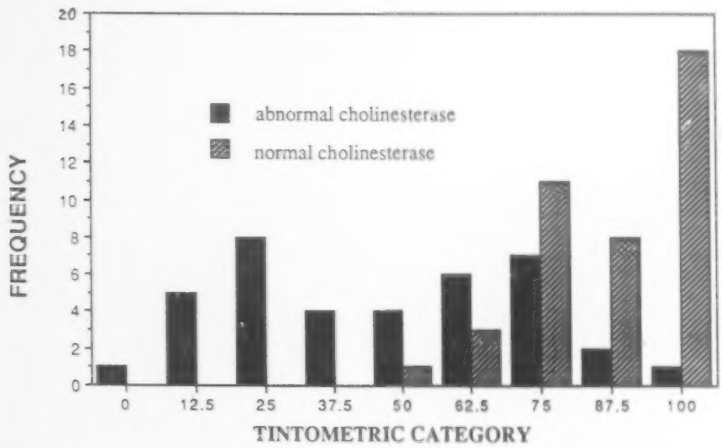


FIGURE 1—Frequency of normal and abnormal cholinesterase results (according to reference assay) in each tintometric category (graphed as percentage of control activity).

phate insecticides for 2 weeks; then retest to assess recovery.

- ≤50% of control cholinesterase activity: serious overexposure; repeat test. If results are confirmed, suspend subject from all work with insecticides. If subject is ill, or if results are less than 24% of control activity, arrange medical examination.

The results of the reference assay for erythrocyte cholinesterase were used for evaluating the tintometric kit. The assay was conducted as previously described:[7] 46 healthy, unexposed male health workers were sampled to establish a lower normal 90% confidence limit of 3.7 IU/minute/mL of blood (mean = 4.6 IU; SD = 0.54).

Results

The number of workers in each tintometric (whole blood cholinesterase) category and the mean erythrocyte cholinesterase activity (reference assay) among workers in each tintometric category are presented in Table 1. In a previous comparison of the tintometric method, a similar pattern of increasing mean erythrocyte cholinesterase was seen as the tintometric categories increased.[11] Thirty-eight of the 79 workers (48%) had erythrocyte cholinesterase activity (reference assay) below the lower normal limit of 3.7 IU. Thirty-two of 79 (41%) had low tintometric cholinesterase activity, using tintometric category 75% as the criterion for abnormal. However, in the critical tintometric categories from 75% to 50% of control activity, there was considerable overlap of workers with normal and abnormal erythrocyte cholinesterase activity, as determined by the reference assay (see Table 1 and Figure 1). In addition, although all workers with tintometric activity of less than 50% also had abnormal activity according to the reference assay, 3 of the 29 workers (10%) with tintometric activity of 87.5% or 100% of control activity had low activity according to the reference assay.

The sensitivity, specificity, and predictive value of the tintometric kit were calculated considering as abnormal a tintometric category of 75% or below (Table 2), as recommended by the manufacturer. In this highly exposed population, predictive value positive was only 70%. When tintometric categories 62.5% and 50% were used as criteria for abnormal and the screening parameters were recalculated, sensitivity declined markedly and specificity improved dramatically. Unfortunately, for any of the three tintometric categories (75%, 62.5%, and 50% of control activity), either sensitivity or specificity was less than 75%, and predictive value positive or predictive value negative was less than 80%.

Discussion

Although the overall correlation between the reference assay and the tintometric tests was good, either sensitivity or specificity was less than 75% for each of the three critical tintometric categories commonly used to define abnormal (from 75% to 50% of the tintometric control). In addition, although there is good agreement between the two methods at tintometric values of less than 50% of control activity, in tintometric categories of 87.5% and 100%, 3 of 29 samples were low by the reference assay. According to the tintometric kit's manufacturers, these samples should all be normal.[12] These results are consistent with our previous observation that the tintometric kit substantially

TABLE 2—Sensitivity, Specificity, and Predictive Value of the Tintometric Kit, with the Use of Different Tintometric Categories to Define Abnormal: Crop-Dusting Airports, Nicaragua (n = 79)

Tintometric Abnormal	True-Positive	False-Positive	True-Negative	False-Negative	Sensitivity, %	Specificity, %	Predictive Value Positive, %	Predictive Value Negative, %
75	38	15	26	3	92	63	70	90
62.5	28	4	37	10	74	90	88	79
50	22	1	40	16	58	98	96	71

overestimated cholinesterase activity during recovery from acute poisoning.[7]

Although there is a disturbing lack of agreement between the tintometric cholinesterase assay and the Ellman assay, the results may actually overestimate the accuracy to that kit. The intraindividual variance of erythrocyte cholinesterase over time is much less than the interindividual variance (used in this study to establish a normal range). Because of the wide range of normal values in different unexposed individuals, it is recommended that a preexposure baseline cholinesterase activity be established for each pesticide-exposed worker for subsequent comparison. Any organophosphate-exposed worker whose erythrocyte cholinesterase falls to 70% of baseline should be removed from exposure.[3,13,14] Unfortunately, the tintometric assay does not have sufficient precision to establish such a baseline reliably. There might have been both more false-positive and more false-negative results if preexposure reference baseline levels had been available for these workers. In addition, it should be noted that this was a heavily exposed population with a very high prevalence of workers with depressed cholinesterase according to the reference standard. It is well known that predictive value depends on prevalence in the sample population.[15] In a population with lower prevalence, predictive value positive would decrease.

Caution should be used in interpreting results of this useful biomarker of organophosphate exposure, especially when screening populations with low prevalence of overexposure, unless appropriate validation and rigorous quality control guarantee accuracy. □

Acknowledgments

This study was supported by the Nicaraguan Ministry of Health, CARE Nicaragua, American Friends Service Committee, grant 1 K01 OH00123-01 from the National Institute for Occupational Safety and Health of the Centers for Disease Control and Prevention, and grant P30 ES00928 from the National Institute of Environmental Health Sciences.

We appreciate the help of Mario Cordón, Carmen Ibania Kuan, Douglas L. Murray, Mario Pais Niño, Donald Roque, Feliciano Pacheco Antón, and Wiron Valadares.

References

1. Jeyaratnam J. Acute pesticide poisoning: a major global health problem. *World Health Stat Q.* 1990;43:139–144.
2. *Public Health Impact of Pesticides Used in Agriculture.* Geneva, Switzerland: World Health Organization; 1990.
3. Criteria for a recommended standard. Occupational exposure during the manufacture and formulation of pesticides. Washington, DC: National Institute for Occupational Safety and Health; 1978:8. DHEW publication 78–174.
4. Edson EF. Blood tests for users of OP insecticides. *World Crops.* 1950;10:49–51.
5. Miller S, Shah MA. Cholinesterase activities of workers exposed to organophosphorus insecticides in Pakistan and Haiti and an evaluation of the tintometric method. *J Environ Sci Health.* 1982;B17(2):125–142.
6. Ellman GL, Courtney KD, Andres V, Featherstone RM. A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochem Pharmacol.* 1961;7:88–95.
7. Magnotti RA, Dowling K, Eberly PE, McConnell R. Field measurement of plasma and erythrocyte cholinesterases. *Clin Chim Acta.* 1988;176:315–332.
8. McConnell R, Cedillo L, Keifer M, Palomo MR. Monitoring organophosphate insecticide exposed workers for cholinesterase depression: new technology for office or field use. *J Occup Med.* 1992;34–37.
9. King ME. Cholinesterase. In: Kaplan LA, Pesce AJ, eds. *Clinical Chemistry: Theory, Analysis, and Correlation.* St. Louis, Mo: CV Mosby; 1989.
10. McConnell R, Pacheco F, Magnotti R. Crop dusters aviation mechanics: high risk for pesticide poisoning. *Am J Public Health.* 1990;80:1236–1239.
11. McConnell R, Cordón M, Magnotti R, Murray D. Hazards of closed pesticide mixing and loading systems: the paradox of protective technology in the Third World. *Br J Ind Med.* 1992;49:615–619.
12. Thomas LC, Chamberlain GJ, Schute C. The rapid field determination of cholinesterase. In: *Colorimetric Chemical Analytical Methods of the Tintometer Company.* 9th ed. New York, NY: John Wiley & Sons; 1980:397–400.
13. Expert committee on the safe use of pesticides in public health: technical report series no. 356. Geneva, Switzerland: World Health Organization; 1967.
14. Epidemiological Studies Laboratory. *Medical Supervision of Pesticide Workers: Guidelines for Physicians (AFC 59122).* Berkeley, Calif: State of California, Department of Health; 1988.
15. Last JM. Epidemiology and health information. In: Last JM, ed. *Public Health and Preventive Medicine.* Norwalk, Conn: Appleton-Century-Crofts; 1986;47–49.

Public Health Then and Now

Editor's Comments: Plurality of Views

In view of the timeliness and seriousness of the issues under discussion, the Editors agreed to publish the following exchange between Howard Waitzkin and Alain Enthoven on the history of managed competition. The proposals for health care reform currently under national discussion span a broad range from "single payer" to the unfettered free market, and we recognize that these different proposals inspire strong feelings and opinions. The American Public Health Association has long been a strong proponent of national health care reform and historically has endorsed the principles of the single-payer position. Although the editors of the *American Journal of Public Health* welcome spirited argument on the issues of health care reform, we would like to emphasize that this journal is open to a plurality of views, and we reject any implication that one approach is more "American" than another. □

Elizabeth Fee

The Strange Career of Managed Competition: From Military Failure to Medical Success?

Howard Waitzkin, MD, PhD

ABSTRACT

Managed competition remains untested as the basis of a national health program. However, key principles of managed competition first emerged in the military. For this study, published works on systems analysis and the planning-programming-budgeting system (PPBS), developed by Alain Enthoven and colleagues at the US Department of Defense during the 1960s, were compared with published presentations of managed competition. The influence of PPBS waned after it generated controversy and opposition. PPBS and managed competition represent similar managerial strategies of policy reform. Although the origin of managed competition in failed military policy does not ensure failure in the medical arena, this history also does not augur success. (*Am J Public Health.* 1994;84:482–489)

> Beyond its uniqueness and eclecticism, I would like to say that the art of weapon systems analysis, like the art of medicine, should be based on scientific method, using that term in its broadest sense.[1]
>
> —*Alain C. Enthoven, 1963*

Since managed competition as the basis of a national health program remains untested anywhere in the world, by what yardstick should it be judged?

Current debates do not recognize that key principles of managed competition arose historically in a completely different, nonmedical arena: the military. These principles briefly became influential at the US Department of Defense and certain other parts of the federal government. However, this approach to policy-making eventually fell into disfavor because of both its failure to achieve policy goals and the widespread opposition that it generated in the military sector, other administrative branches of government, Congress, and the general public.

The origins of managed competition in military policy help clarify what can be expected from a similar managerial strategy of health care reform. Through a critical review of published works, this article traces the history of managed competition, compares complementary themes in the military and medical arenas, and presents implications for future decisions about a US national health program.

Origins of Managed Competition in Military Policy Analysis

The conceptual framework of managed competition originated in the military situation that confronted the United States during the Cold War of the late 1950s and early 1960s. Professor Alain Enthoven, an economist and the princi-

The author is with the Department of Medicine and School of Social Sciences, University of California–Irvine.

Requests for reprints should be sent to Howard Waitzkin, MD, PhD, University of California–Irvine, North Orange County Community Clinic, 300 W Romneya Dr, Anaheim, CA 92801.

pal intellectual architect of managed competition, was a driving force behind systems analysis at the Pentagon between 1961 and 1969, and his conceptual approach to health care owes much to the analytic work that he and his colleagues carried out during that period. Military systems analysis and managed competition in health care represent similar managerial approaches to policy reform.

After John F. Kennedy became president in 1961, he appointed Robert McNamara, chief executive officer of the Ford Motor Company, as secretary of defense. An advocate of strong management, McNamara selected Charles Hitch, an economist based at the RAND Corporation, as assistant secretary of defense. While at RAND during the 1940s and 1950s, Hitch had spearheaded the Economics Division, which was funded mainly by the Air Force and whose purpose was to apply economic analysis to choices of weapon systems and strategies; this work at RAND culminated in an influential book that spelled out how economic analysis could guide defense policy decisions.[2] Hitch brought with him to the Pentagon several colleagues from RAND, including Enthoven, who helped establish the Office of Systems Analysis and later also rose to become assistant secretary of defense. Enthoven presented the accomplishments of this office in a series of articles appearing in military and economics journals during the 1960s, as well as in two books.[3,4]

At the Pentagon, Enthoven and his coworkers emphasized the planning-programming-budgeting system (PPBS) and cost-benefit analysis.[4] PPBS aimed to develop "plain statements" of national purposes in defense programs. Based on such formulations, each defense program could be evaluated and compared with alternative programs aimed at achieving the same purposes. In considering military goals, this approach tried to express costs and benefits quantitatively (see photo). An independent analytical staff, composed mostly of civilian specialists such as economists, prepared technical reports comparing the alternatives that the secretary of defense and other key officials used in decision making. The independence of this staff was viewed as essential to balance the narrow interests of military officers, defense contractors, and politicians who sought to influence decisions.

As systems analysis gained prominence, several policy decisions reflected the impact of PPBS. These decisions concerned deployment of troops in the North Atlantic Treaty Organization; development and proliferation of nuclear weapons; and initiation of new bombers, missiles, and other weapon systems.[4] Because of the apparent rationality and high-level sponsorship of this approach, President Lyndon Johnson in 1965 mandated the gradual implementation of PPBS throughout the federal government.[5]

At various points in the Vietnam War, systems analysts sought to train military officers through hypothetical exercises such as the evaluation of resupply operations by ground versus air.[6] According to Enthoven, PPBS also contributed to the war effort by evaluating troop deployment plans, developing alternatives to "body counts" as measures of success in the war effort, estimating aircraft attrition, and measuring progress in the Vietnam "pacification" program. But despite its attempt to reform military policies through a strategy of strong and rational management, systems analysis generated controversy and opposition both within and outside government, and its influence began to wane as US participation in the Vietnam War increased during the late 1960s.

In a retrospective account, Enthoven and K. Wayne Smith, a colleague at the Pentagon, acknowledge this opposition.[4(pp267–308)] Administrators of other cabinet-level departments did not consistently implement President Johnson's mandate. Professional military leaders, defense contractors, and civilian analysts at the Pentagon were antagonistic toward PPBS. Politicians in Congress grew impatient with quantitative analyses of defense policies (see photo of Gardiner L. Tucker).[4(pp309–313)] Antiwar activists argued that cost-benefit analysis led to inappropriate policies because it clouded the ethical dimensions of important decisions, especially those about nuclear weapons and Vietnam. But despite its claimed advantages in cost controls, PPBS did not prevent or modulate a rapid increase in military expenditures, which remained unsurpassed until the mid-1980s (Figure 1).[7(p336)] This rate of increase in military expenditures during the 1960s reflected the pressures of the Vietnam War and concerns about nuclear conflict, and PPBS was not in a position to control overall costs in the context of these broad political forces.

DR. ALAIN ENTHOVEN
He tells the military what's right.

Alain C. Enthoven lecturing on systems analysis in the military, 1963. On the chalkboard, the vertical axis is labeled "targets destroyed" and the horizontal axis is labeled "$." Reprinted with permission from Time, Inc.[50]

Thus, while Enthoven and his associates continued to defend PPBS, they acknowledged that their principles had little impact on military policies during the late 1960s. Accordingly, Enthoven's retrospective view conveys disappointment and failure:

> The Systems Analysis office did not have a prominent, much less a crucial, role in the Vietnam war. . . . It had no policy role in determining the overall totals of men to send to Vietnam, or in figuring out what they should do when they got there. . . . If we make no other point in this book, we want it to be clear that the full value of systematic analysis in making decisions on the conduct of a war has yet to be tested.[4(p270)]

Already in eclipse, PPBS continued to lose prominence in policy-making, and its impact in other spheres of government remained minimal. As Enthoven and Smith point out, President Richard Nixon promised to "root out the

Gardiner L. Tucker, Enthoven's successor as head of the Office of Systems Analysis at the US Department of Defense, 1970. The caption refers to congressional criticism of systems analysis. Department of Defense file photo that appeared in *Business Week*.[51]

whiz kid approach at the Pentagon."[4(p333)] Although the Office of Systems Analysis survived, its activities became much more circumscribed. Temporary decreases in military spending during the mid-1970s mostly reflected the scaling back of operations in Southeast Asia rather than the impact of PPBS, which by that time contributed in a very minor way to budgeting at the Pentagon. PPBS did not enjoy a comeback under the administration of Jimmy Carter, and it remained uninfluential during the peacetime military buildup of the Reagan and Bush administrations.

After a brief excursion into private industry between 1969 and 1973 as vice president and then president of Litton Industries, a major military contractor, Enthoven then joined the faculty of Stanford University in 1973 as a professor of both management and health care economics.[5] There his work on health policy incorporated the principles of management and systems analysis that he had previously advocated at the Pentagon.

By 1977, only 4 years after leaving the defense sector, Enthoven offered to the Carter administration a proposal for a "Consumer Choice Health Plan," based on regulated competition in the private sector. This proposal built in part on prior initiatives by Paul Ellwood for a national health maintenance strategy and by Scott Fleming for structured competition within the private sector.[9(pp65–67)] Although Carter rejected the plan, Enthoven soon afterward published the proposal in the medical literature[10] and in a separate monograph.[11] In this early work, Enthoven presents the basic conceptual structure of all subsequent proposals for managed competition. As discussed later, this proposal for health care reform contains important concepts from the military policy work that Enthoven had spearheaded a few years earlier at the Pentagon.

During the 1980s, Enthoven collaborated with Ellwood, other proponents of health maintenance organizations (HMOs), corporate executives, and officers of private insurance companies in refining the proposal. An emphasis on managed competition arose during the mid-1980s in response to concerns raised by economists and business leaders that the original proposal conveyed free-market assumptions requiring modification through closer "management" of the program.[12] The new name also conveyed a message of managerial control that proved attractive to business leaders.

After publication of a revised proposal in 1989,[13] the coalition supporting managed competition broadened to include officials of the largest US private insurance companies, which were diversifying into managed care. These business leaders met regularly with Enthoven and other proponents of managed competition at Ellwood's Wyoming home as part of the so-called Jackson Hole group. The managed care sector of the private insurance industry provided major funding for this group, as well as financial and logistic support for Bill Clinton's presidential campaign and consultation for the Presidential Health Care Task Force.[14,15]

Most recent proposals for managed competition—including that of the Clinton administration—have emerged from this intellectual tradition. These proposals use fixed premium contributions by employers and government to encourage enrollees to choose less expensive plans; the proposals also use price competition among health plans to help control fees paid to doctors and hospitals.[16,17] Certain proposals have suggested modifications in the conceptual structure outlined by Enthoven and his colleagues. For instance, although managed competition traditionally has encouraged employer-sponsored plans with participation by private insurance companies, other proposals have separated employment from insurance through the creation of a single, tax-financed, globally budgeted public fund, which would contract with private plans for a minimum benefit package.[18–20] All managed competition proposals, however, incorporate concepts initiated by Enthoven and his colleagues, and all of them call for large-scale changes in the ways medicine is practiced and choices are made by physicians and consumers.

Complementary Themes in Military and Health Policy

As has been seen, the theoretical advantages of strong management techniques, as advocated by Enthoven and his colleagues, did not lead to enduring improvements in defense policy. This failure derived not only from the self-interested opposition of military officers and members of Congress, but also from the limitations of PPBS in addressing the political and ethical dimensions of major public policy decisions. Because the theory of managed competition in health care, which seems to be heading for its possible first test, contains unmistakable elements of PPBS, the limitations and complementary themes in these two approaches to public policy merit closer critical scrutiny than they have received thus far. Table 1 summarizes these themes.

Distrust of Professionals

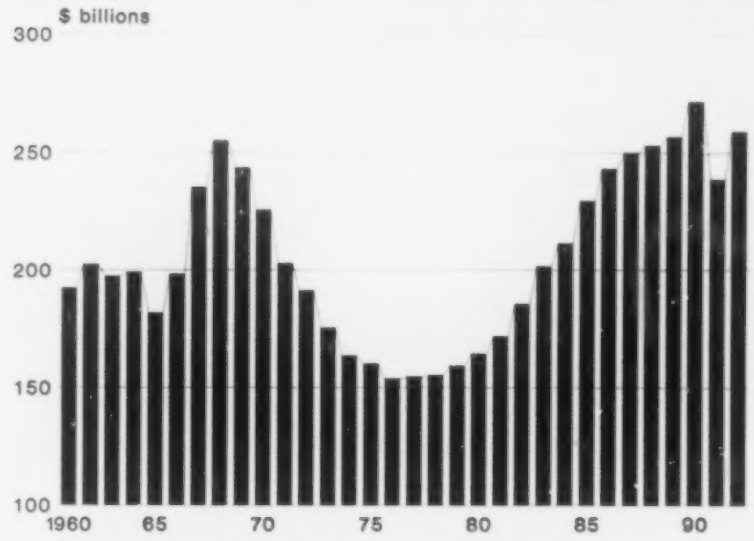
PPBS argued that professional military officers held vested interests that led them to advocate ever-expanding investment and program development in their own narrow areas of expertise. For instance, the chiefs of the army, navy, and air force promoted new weapon systems with little critical appraisal of those systems' contribution to the national interest and without coordination among the services. While expressing formal respect for the professional competence of military officers, Enthoven and other proponents of PPBS argued that defense policy must counterbalance the vested interests of military brass through technical analyses of costs and benefits performed by civilian personnel. Independent analysts were needed to foster informed and cost-conscious defense policy decisions.[4(pp42–44)]

In managed competition, distrust of professionals manifests itself in a very critical assessment of physicians' vested interests. According to Enthoven, physicians function as a guild, which reduces the ability of market forces to control costs or encourage quality. From this perspective, the guild structure permits free choice of doctor by patient, free choice of prescription by doctor, direct negotiation between doctor and patient regarding fees, payment based on fee-for-service, solo practice, and professional control over medical licensure.[9,10] Managed competition, on the other hand, calls for concerted action by independent analysts and purchasers to curb decision-making power based on physicians' professional self-interest.

Trust of Managers

In arguing for the role of civilian analysts at the Department of Defense, Enthoven and his colleagues placed confidence in "active management at the top." In this model, implemented at the Pentagon, the secretary of defense received proposals about programs and weapon systems from military officers through the joint chiefs of staff. But independent analysts then subjected these proposals to rigorous analysis of costs and benefits in light of precise statements of the national interest. Memoranda prepared for the president presented conclusions of these analyses, thus encouraging more informed and objective decision making.[4(pp73–116)] Throughout this process, it was assumed, civilian managers would prove less guided by vested interests than would the professional brass.

Likewise, the wisdom of managers mostly based outside the medical profession has emerged as a basic tenet of managed competition. Under this approach, the locus of decision making moves from the medical guild to a cadre of managers, who administer the organizational "sponsors" of group health plans. (Under recent proposals, these sponsors are called "health insurance purchasing cooperatives" or "health alliances.") Incentives encourage physicians to organize themselves into competing economic units. Managers of the sponsors, which could include employers, labor-management health and welfare trusts, the federal Medicare system, state governments, and other large organizational entities, make prudent choices among competing plans on behalf of individuals and families. Guided by managerial principles, the sponsors serve as "active, intelligent, collective agents on the demand side who structure and adjust the market in a continuing, but never completely successful, effort to overcome its tendencies to failure."[9(p75)]



Note. Military expenditures increased rapidly during the 1960s, including the early years of the Vietnam War, when PPBS, which encouraged cost-conscious policy decisions, was in effect at the US Department of Defense.
Source. The data are from *Statistical Abstract of the United States* (Table 525, p 336, 1992 edition; data not given for 1962).[7]

FIGURE 1—Growth in US military expenditures, in constant (1982) dollars, from 1960 to 1992.

TABLE 1—Complementary Themes in Military Planning-Programming-Budgeting System (PPBS) and Managed Competition in Health Care

Theme	Focus of Theme	
	PPBS	Managed Competition
Distrust of professionals	Military brass	Medical guild
Trust of managers	Independent analysts	Managers of organizational sponsors
Choice among competing alternatives	Weapon systems, military strategies	Organized health plans
Scientific method	Cost-benefit analysis in case studies	Cost-benefit analysis in case studies
"Tools" for managers	Cost-benefit analysis, 5-year defense plan, draft presidential memos, development concept papers	Techniques to prevent "market failure": pricing, cost-benefit analysis, annual enrollment, quality assurance, subsidy management
Incrementalism with strains	Military officers, corporate contractors, members of Congress	Physicians, middle-class consumers, private insurance companies not equipped for managed care
Cost analysis but not necessarily cost reduction	No budget ceiling for military expenditures	No budget ceiling for health care expenditures

TABLE 2—Example of the Planning-Programming-Budgeting System as Applied to Nuclear Weapons: Comparative Effectiveness of Two Hypothetical Missile Pay Loads,[a] by Number of Targets Destroyed

Type of Target Destroyed	Ten 50-Kiloton Warheads Totaling One-Half Megaton	One 10-Megaton Warhead
Airfield	10.0	1.0
Hard missile silos	1.2–1.7[b]	1.0
Cities of 100 000 population	3.5	1.0
Cities of 500 000 population	0.7	1.0
Cities of 2 000 000 population	0.5	0.6

Source. Reprinted with permission from Enthoven and Smith.[4(p182 Table II)]
[a]Both assumed to be reliably delivered.
[b]Variation depends on target hardness, delivery errors, and number of warheads allocated to each silo.

This view also calls for a reduction in decision-making discretion for individuals and families as consumers of services. Because of technical complexity, managers assume a paternalistic role in making rational decisions on behalf of consumers. As Enthoven writes: "This isn't a market in which the invisible hand will do the job. Some visible hands, which I call sponsors, must manage the demand side to make the market achieve desirable results."[9(p113)] Managerial input would shape not only the range of health plans offered to consumers but also the basic benefits package. Like PPBS, managed competition places faith in the ability of technically oriented managers to make decisions that are independent of any particular interest group. Thus, as decision making moves to high-level managers, managed competition is expected to foster administrative authority wielded in an enlightened effort to reconcile the interests of physicians, patients, business, and government.

Choice among Competing Alternatives

PPBS sought to analyze the marginal costs and benefits of competing alternatives to similar policy goals so that high- level managers could make informed choices. In the case of nuclear weapons, Enthoven and his colleagues tabulated hypothetical data on the projected number of enemy cities destroyed by multiple warheads, each containing a relatively small nuclear payload, as opposed to fewer warheads, each with a much larger payload (Table 2). Through these tabulations, it became apparent that multiple warheads with smaller payloads would prove more cost-effective since more cities with larger cumulative populations could be destroyed for a given level of investment in nuclear weapons.[4(pp165–195)] In the context of Cold War pressures, this analysis coincided with one of the largest overall buildups of nuclear weapons in history, and the ethical dimensions of such policy alternatives received very little attention.

By a similar logic, managed competition encourages the development of competing health plans, among which managerial sponsors would choose on behalf of consumers. In making informed choices, managers would consider data on the relative costs and benefits of competing alternatives. In some instances, choice would also be offered to consumers, but they would have to pay more for health plans that provide benefits beyond the basic minimum package. The main locus of choice, however, will reside with the managers of large organizations that contract with competing health plans. From this viewpoint, the theoretical advantages of managerial choice are seen as superior to alternative policies such as a single-payer system, which controls costs through monopsony financing while permitting substantial choice for individual patients and doctors.

Scientific Method

At the Pentagon, PPBS introduced quantitative, cost-oriented analyses into policy deliberations. This approach conveyed scientific rationality, as opposed to prior procedures that were based on subjective opinion and the power positions of military and congressional leaders. A frequently used method of analysis was the case study, such as the study of nuclear warheads and payloads cited above, in which the costs and benefits of alternative weapon systems or military strategies were compared. The complexity and esoteric features of such studies tended to make them inscrutable to military officers and congressional representatives unschooled in economics.[4(p62)] For policymakers, decisions often involved nonquantifiable considerations, mostly value judgments, which proponents of PPBS referred to as "imponderables."[2] Further, PPBS required assumptions that simplified complex realities of military operations. Analysts at the Pentagon often used theoretical or hypothetical exercises that, despite their sophistication, proved much less useful than hoped in guiding concrete operational decisions, such as the deployment and resupply of equipment in Vietnam.[6]

The aura of scientific method also pervades arguments on behalf of managed competition. Again, rather than calling for more powerful scientific methods, such as random controlled trials on a regional or state level, to evaluate managed competition as opposed to single-payer or other policies, Enthoven uses case studies selectively to justify key claims. For instance, he uses the Kaiser system as a case study to demonstrate the cost-saving potential of HMOs.[9(pp42–58),10(pp68–69)] In this case study, however, Enthoven downplays prior and subsequent evidence showing that, in contrast to Kaiser, most managed care organizations, such as independent practice associations and preferred provider organizations, have not shown consistent cost savings.[21–23] He and other proponents of managed competition also cite certain cost reductions noted for prepaid group practices within the RAND health care experiment, but they do not emphasize that this random controlled trial did not evaluate managed competition per se.[13] In his case study of Kaiser, Enthoven also does not address data that show similar or lower levels of patient satisfaction in HMOs than in fee-for-service settings.[24–26] Similarly, he uses Canada as a case study and alludes to the work of economist Robert Evans,[27] which showed increases in payments to physicians under Canada's national health program; however, while Enthoven cites a small facet of Evans' work to argue against a single-payer

system,[9(p36)] he does not refer to the larger corpus of Evans' or other investigators' work that has yielded favorable evaluations of the Canadian system.[28–31]

The science of managed competition therefore remains based in theories of market forces and selective application of case studies, rather than in evaluation using more powerful methods available in health services research. Like PPBS before it, managed competition thus becomes a set of scientifically oriented principles to be implemented without prior scientific test.

"Tools" for Managers

New analytic "tools" became corresponding elements of both PPBS and managed competition. In PPBS, chief among these tools were the hypothetical case studies that considered costs and benefits of alternative policies, especially regarding weapon systems. Other tools introduced by Enthoven included a 5-year defense plan, draft presidential memoranda, and development concept papers that presented competing policy options for executive decision.[4(pp48–60)] Efficacious military policies depended on the disciplined use of these tools by systems analysts and their superiors in top management.

The tools of managed competition, again envisioned mainly by Enthoven, are techniques that managers of organizational sponsors can use to counteract "market failure." These tools include methods of pricing, construction of standardized benefit packages based on analysis of projected costs and benefits, annual enrollment and disenrollment procedures, quality assurance methods, management of subsidies, and various "pro-competitive actions" that restrict the economic power of health care providers.[9(pp98–118)] A manager's task would be to use such tools to reduce the impact of market failures, such as continuing access barriers, selection of low-risk enrollees, and cost overruns. Practical impediments to using these managerial tools remain understated, as was the case with similar tools at the Pentagon.

Incrementalism with Strains

Proponents of PPBS acknowledged that their methods implied drastic changes in military policy-making. Incrementalism, the only route ideologically acceptable in the "free world," comprised the process by which advocates believed such changes would occur.[32,33] Career officers, corporations that received military contracts, and people in Congress who supported the traditional way of doing business experienced the inevitable "strains" of incremental change. Military and congressional resistance to such strains eventually contributed to the decline of PPBS as a managerial technique in government. Again, PPBS proved unrealistic in its assessment of US political realities as it underestimated the difficulties of achieving fundamental policy shifts through slow, incremental processes.

Under managed competition, the guild of medical professionals, patients unaccustomed to making cost-conscious decisions, and private insurance companies unequipped for managed care would experience the main strains of change, which again would occur through an incremental process. Arguing that "incrementalism is a fundamental law of behavior in democratic governments," Enthoven advocates a coordinated strategy to "break up the medical guild."[9(pp119–124)] This strategy includes legislation that permits selective contracting with preferred providers, outlaws boycotts by physicians, prohibits price fixing, and restrains other "concerted refusals to deal" by medical professionals. To achieve the widest impact, Enthoven suggests that incremental reforms start with the "educated middle class," who are already accustomed to cost-conscious choices. In discussing incrementalism, however, Enthoven again tends to downplay opponents' ability to undermine processes of slow, incremental change. Moreover, he and his colleagues also discount an alternative view of social change: that democratic societies are capable of making fundamental changes rapidly and successfully, as happened when the United States implemented Social Security in 1935 or when other countries enacted single-payer national health programs.

Cost Analysis but Not Necessarily Cost Reduction

While it analyzed competing alternatives in terms of costs and benefits, PPBS never promised to reduce costs overall. Aiming to achieve the "fullest value per dollar spent,"[5] its advocates viewed PPBS as a better way to manage that was compatible with higher, lower, or unchanged costs. For instance, Hitch noted that cost-benefit analysis remained neutral regarding the unit cost of a particular weapon system.[34] In arguing for explicit consideration of both costs and needs, Enthoven opposed arbitrary budget ceilings that would restrain overall increases in military expenditures.[4(p203)] As previously noted, PPBS coincided with an unprecedented rate of growth in military spending (Figure 1).

Similarly, proponents of managed competition are careful to note that it will not necessarily lower health care costs. Again opposing a cap on overall expenditures, Enthoven favors, on theoretical grounds, market forces that encourage cost-conscious choices, with a goal of maximizing value for money. He and his colleagues predict short-term increases in health care expenditures, as the costs of new layers of administration are added to those of expanded services for the currently uninsured. While longer-term savings may result from managerial decision making, Enthoven and his colleagues acknowledge that such savings remain largely hypothetical and will take years of organizational change to achieve.[9(pp119–135)]

Conclusion

Managed competition has had a strange career. Although this approach remains untested as the organizing principle of any other country's national health program, its intellectual progenitor enjoyed a brief heyday in military policy. By developing and promoting PPBS at the Pentagon, Enthoven and his colleagues worked through many of the economic and analytic procedures later resurrected in the theory of managed competition. In lieu of other yardsticks to evaluate the claims of managed competition, the history of PPBS provides a sense of what may lie ahead in health care.

PPBS failed as a coherent approach to military policy—a failure that deserves attention as managed competition, a similar managerial strategy of reform, gains ascendancy in health policy. PPBS unrealistically gauged the political and ethical dimensions of major public policies. Generating widespread opposition among military professionals, corporate contractors, members of Congress, and peace activists, this managerial strategy of reform exerted little long-term impact on military operations and expenditures or on other spheres of

public administration. Similarly, professional opposition has undermined managed care initiatives in several European countries, both earlier in this century and more recently[35,36]; thus, it is unlikely that medical professionals and their organizations in the United States, seldom lacking in ingenuity in the past, will accept without resistance the administrative control that managed competition here promises to enact. Moreover, the touted advantages of PPBS and managed competition in scientific method, expanded choice, and cost-conscious decisions by high-level managers have not achieved prior successes in policymaking at the national, regional, or state level. Incrementalism with strains, seen as the process of change under both PPBS and managed competition, reflects an ideological, or at least debatable, view of historical processes. The analytic tools of PPBS and managed competition have not lowered overall costs in either the military or medical sector, nor were they intended to do so. Although the origin of managed competition in failed military policy does not guarantee failure in the medical arena, this history also does not augur success.

Aside from this history, both supporters and opponents have raised major concerns about managed competition.[37–41] Demographic limitations would restrict its impact since about 30% of the US population live outside metropolitan areas that could support three or more competing managed care plans. Whether managed competition could control costs remains unclear; states with the most extensive managed care programs have shown costs to be as high as or higher than elsewhere.[42–44] Moreover, administrative costs, already more than 20% of overall health care expenditures,[45] likely would increase still further since managed care is administratively intensive and new organizational sponsors would introduce additional managerial layers. Despite a laudable intent to use research on effectiveness and outcomes to define the uniform minimum benefits package and to assess quality of care, such research has produced verified data about only a small number of medical conditions and procedures. Although "outcomes research" may be construed as increasing professional accountability, proposals for managed competition have not clearly described who would carry out such research or whether access to data would remain within the public domain and in usable form. How practice guidelines would achieve accountability within for-profit entities that compete on the basis of price, while not alienating providers and patients, remains unclear.

Several practical concerns have also arisen regarding the acceptability of managed competition to providers and consumers.[15,36] While expanding the decision-making power of large insurance companies that already have entered the managed care field, managed competition probably will reduce consumers' freedom to choose practitioners, and micromanagement of clinical decisions likely will increase. Because the ability to buy additional coverage beyond the basic benefits package will depend on income, this provision will perpetuate unequal, multitiered coverage. Whether managed competition will succeed in curbing insurance companies' selection or exclusion of patients by risk of costly illness remains in doubt. Managed competition likely will create higher out-of-pocket payments and taxes for a substantial part of the population who currently are insured. Several polls have shown less public support for managed competition than for other alternatives.[46] Evidence that managed competition would solve the access problem while controlling costs remains less convincing than evidence that exists in support of a single-payer option.[47,48] For instance, the US Congressional Budget Office has reported that prospects for cost containment under managed competition are uncertain at best and that the only countries that have successfully contained costs are those using single-payer financing.[21]

The powerful coalition built up around managed competition may succeed in enacting this policy. Although it may address some of the concerns raised about managed competition, the Clinton team is unlikely to change the basic structure of the proposal. This reluctance to consider other options seriously may stem partly from the support that the Clinton campaign received from the managed care sector of the private insurance industry,[14,15] as well as from a perception that simpler and more popular options, including a single-payer approach, are unlikely to pass in Congress.

Failure to achieve a workable national health program would generate great disappointment as well as financial waste. Some analysts believe that failure of managed competition is a necessary step toward adoption of a simpler approach such as a single-payer option. A less sanguine view holds that the United States may get only one chance to establish a workable program during this decade and that failure would lead to retrenchment, cutbacks, and a return to the present paradox of pervasive access barriers coupled with high costs.

Ultimately, should market principles of competition, derived largely from failed attempts at managerial reform of military policy, be accepted as cornerstones of national health policy? It is unnecessary to expound on the ways in which health care differs from military procurement. But having emerged historically from military policy analysis, managed competition contains assumptions about the superiority of managerial decision making and discretion, even as they are applied to the micromanagement of clinical processes. Such assumptions about managerial wisdom may lead medicine down a path inconsistent with the aspirations of many health workers and patients. Rather than Enthoven's equation of the art of weapon systems analysis and the art of medicine, an alternative vision might be Wendell Berry's:

> Rats and roaches live by competition under the law of supply and demand; it is the privilege of human beings to live under the laws of justice and mercy. It is impossible not to notice how little the proponents of the ideal of competition have to say about honesty, which is the fundamental economic virtue, and how *very* little they have to say about community, compassion, and mutual help.[49(p135)] □

Acknowledgments

This study was supported in part by grants from the Poverty and Race Research Action Council (Washington, DC) and the US Health Resources and Services Administration (PE 19154).

I thank Stephany Borges, Carolyn Clancy, David Himmelstein, Don Light, Vicente Navarro, Steffie Woolhandler, members of the Health Services Research Seminar at the University of California–Irvine, Elizabeth Fee, and anonymous Journal reviewers for helpful comments on an earlier draft of this article. I am grateful to Michelle Stone and Wendy Wagner for bibliographic assistance.

References

1. Enthoven AC. Choosing strategies and selecting weapon systems (address before the Naval War College, Newport, RI, June 6, 1963). In: Tucker SA, ed. *A Modern Design for Defense Decision: A McNamara-Hitch-Enthoven Anthology.* Washington, DC: Industrial College of the Armed Forces; 1966.

2. Hitch CJ, McKean R. *The Economics of Defense in the Nuclear Age*. Cambridge, Mass: Harvard University Press; 1960.
3. Tucker SA, ed. *A Modern Design for Defense Decision: A McNamara-Hitch-Enthoven Anthology*. Washington, DC: Industrial College of the Armed Forces; 1966.
4. Enthoven AC, Smith KW. *How Much Is Enough? Shaping the Defense Program, 1961–1969*. New York, NY: Harper & Row; 1971.
5. Johnson LB. Introduction of new government-wide planning and budgeting system (statement by the president to members of the Cabinet and heads of agencies, August 25, 1965). In: Tucker SA, ed. *A Modern Design for Defense Decision: A McNamara-Hitch-Enthoven Anthology*. Washington, DC: Industrial College of the Armed Forces; 1966.
6. Snyder WP. An illustrative example of systems analysis (prepared for instructional purposes, US Military Academy, spring 1964). In: Tucker SA, ed. *A Modern Design for Defense Decision: A McNamara-Hitch-Enthoven Anthology*. Washington, DC: Industrial College of the Armed Forces; 1966.
7. *Statistical Abstract of the United States*. Washington, DC: US Bureau of the Census; 1992.
8. *Who's Who in America*. Chicago, Ill: Marquis; 1992.
9. Enthoven AC. *Theory and Practice of Managed Competition in Health Care Finance*. Amsterdam, the Netherlands: North-Holland; 1988.
10. Enthoven AC. Consumer-choice health plan. *N Engl J Med*. 1978;298:650–658, 709–720.
11. Enthoven AC. *Health Plan: The Only Practical Solution to the Soaring Cost of Medical Care*. Reading, Mass: Addison-Wesley; 1980.
12. Enthoven AC. Managed competition in health care and the unfinished agenda. *Health Care Financ Rev*. 1986;(suppl): 105–119.
13. Enthoven AC, Kronick R. A consumer-choice health plan for the 1990s. *N Engl J Med*. 1989;320:29–37, 94–101.
14. Shapiro B. Zoe Baird. *Los Angeles Times*. January 7, 1993:B11.
15. US health reforms: clichés, cost, and Mrs C. *Lancet*. 1993;341:791–792.
16. Ellwood P, Enthoven A, Etheredge L. The Jackson Hole initiatives for a twenty-first century American health care system. *Health Econ*. 1992;1:149–168.
17. Rice T, Brown ER, Wyn R. Holes in the Jackson Hole approach to health care reform. *JAMA*. 1993;270:1357–1362.
18. Garamendi J. *California Health Care in the 21st Century: A Vision for Reform*. Sacramento, Calif: California Department of Insurance; 1992.
19. Starr P. *The Logic of Health-Care Reform*. Knoxville, Tenn: Whittle Direct Books; 1992.
20. Starr P, Zelman WA. A bridge to compromise: competition under a budget. *Health Aff (Millwood)*. 1993;12(suppl): 7–23.
21. Langwell KM, Staines VS, Gordon N. *The Effects of Managed Care on Use and Costs of Health Services*. Washington, DC: Congressional Budget Office; 1992.
22. Gabel JR, Rice T. Is managed competition a field of dreams? *J Am Health Policy*. 1993;3:19–24.
23. Glaser WA. The United States needs a health system like other countries. *JAMA*. 1993;270:980–984.
24. Davies AR, Ware JE Jr, Brook RH, Peterson JR, Newhouse JP. Consumer acceptance of prepaid and fee-for-service medical care: results from a randomized controlled trial. *Health Serv Res*. 1986;21: 429–452.
25. Health care in crisis: are HMOs the answer? *Consumer Rep*. 1992;57(August): 519–531.
26. Rubin HR, Gandek B, Rogers WH, Kosinski M, McHorney CA, Ware JE Jr. Patients' ratings of outpatient visits in different practice settings. *JAMA*. 1993; 270:835–840.
27. Evans RG. Beyond the medical market place: expenditure, utilization, and pricing of insured health in Canada. In: Andreopoulos S, ed. *National Health Insurance: Can We Learn from Canada?* New York, NY: Wiley; 1975.
28. Evans RG. *Strained Mercy: The Economics of Canadian Health Care*. Toronto, Canada: Butterworths; 1984.
29. Evans RG, Barer ML, Hertzman C. The 20-year experiment: accounting for, explaining, and evaluating health care cost containment in Canada and the United States. *Ann Rev Public Health*. 1991;12: 481–518.
30. Fuchs VR, Hahn JS. How does Canada do it? a comparison of expenditures for physicians' services in the United States and Canada. *N Engl J Med*. 1990;323:884–890.
31. Redelmeier DA, Fuchs VR. Hospital expenditures in the United States and Canada. *N Engl J Med*. 1993;328:772–778.
32. McNamara RS. Managing the Department of Defense (1964). In: Tucker SA, ed. *A Modern Design for Defense Decision: A McNamara-Hitch-Enthoven Anthology*. Washington, DC: Industrial College of the Armed Forces; 1966.
33. Enthoven AC. US defense policy for the 1960s (address at Loyola University, Los Angeles, and at Seattle University, February 1963). In: Tucker SA, ed. *A Modern Design for Defense Decision: A McNamara-Hitch-Enthoven Anthology*. Washington, DC: Industrial College of the Armed Forces; 1966.
34. Hitch CJ. Cost effectiveness (lecture at the University of California, April 1965). In: Tucker SA, ed. *A Modern Design for Defense Decision: A McNamara-Hitch-Enthoven Anthology*. Washington, DC: Industrial College of the Armed Forces; 1966.
35. Abel-Smith B. The rise and decline of the early HMOs: some international experiences. *Milbank Q*. 1988;66:694–719.
36. Glaser WA. The competition vogue and its outcomes. *Lancet*. 1993;341:805–812.
37. Hadley JP, Langwell K. Managed care in the United States: promises, evidence to date and future directions. *Health Policy*. 1991;19:91–118.
38. Iglehart JK. Managed competition. *N Engl J Med*. 1993;328:1208–1212.
39. Kronick R, Goodman DC, Wennberg J, Wagner E. The marketplace in health care reform: the demographic limitations of managed competition. *N Engl J Med*. 1993; 328:148–152.
40. Relman AS. Controlling costs v "managed competition"—would it work? *N Engl J Med*. 1993;328:133–135.
41. Clancy C, Himmelstein DU, Woolhandler S. Questions and answers about managed competition. *Int J Health Serv*. 1993;23: 213–218.
42. Socolar D, Sager A, Hiam P. Competing to death: California's high-risk system. *J Am Health Policy*. 1992;2:45–50.
43. Kosterlitz J. Monopoly medicine? *Natl J*. 1993;25:1746–1749.
44. Access and affordabiiity monitoring project, Boston University School of Public Health. Health costs rose faster in states with higher HMO enrollment during the 1980s. Cited in: Center for National Health Program Studies. *Managed Competition: Interim Update to the National Health Program Chartbook and Slideshow*. Cambridge, Mass: Harvard Medical School/ The Cambridge Hospital; 1993.
45. Woolhandler S, Himmelstein DU. The deteriorating administrative efficiency of the US health care system. *N Engl J Med*. 1991;324:1253–1258.
46. Blendon RJ, Altman DE, Benson JM, Taylor H, James M, Smith M. The implications of the 1992 presidential election for health care reform. *JAMA*. 1992;268:3371–3375.
47. *Canadian Health Insurance: Lessons for the United States*. Washington, DC: General Accounting Office; 1991. GAO publication GAO/HRD–91–90; B–244081.
48. Marmor TR. Commentary on *Canadian Health Insurance: Lessons for the United States*. *Int J Health Serv*. 1993;23:45–62.
49. Berry W. *What Are People For?* San Francisco, Calif: North Point Press; 1990.
50. The whizziest kid. *Time*. June 28, 1963:18.
51. The new mission of systems analysis. *Business Week*. July 4, 1970:70.

A 2623-1

American Journal of Public Health

April 1994, Vol. 84, No. 4 Established 1911

9375
ED KIANKA
HEALTH REPORTER
CONGRESSIONAL INFORMATION SVC
4520 EAST-WEST HIGHWAY
BETHESDA MD 20814

American Public Health Association

Featuring Child Violence and Injury

Editorials and Annotations

537 Editorial: Unintentional Childhood Injury—Where the Buck Should Stop *I. B. Pless*
539 Editorial: Reducing Violence—How Do We Proceed? *K. K. Christoffel*
541 Annotation: Firearm-Related Violence—What We Don't Know Is Killing Us *A. L. Kellermann*

Commentary

543 The Report of the Surgeon General: Preventing Tobacco Use among Young People *M. J. Elders, C. L. Perry, M. P. Eriksen, and G. A. Giovino*

Public Health Policy Forum

548 Editorial: Medicaid Reimbursement—Can We Save Money by Paying Doctors More? *M. Regan*
550 Commentary: Lessons from Medicaid—Improving Access to Office-Based Physician Care for the Low-Income Population *D. Rowland and A. Salganicoff*
553 Medicaid Physician Payment Reform: Using the Medicare Fee Schedule for Medicaid Payments *A. L. Reisinger, D. C. Colby, and A. Schwartz*

Articles

VARIOUS TOPICS

561 Homicide, Handguns, and the Crime Gun Hypothesis: Firearms Used in Fatal Shootings of Law Enforcement Officers, 1980 to 1989 *G. J. Wintemute*
565 The Prevalence and Correlates of Fear of Falling in Elderly Persons Living in the Community *C. L. Arfken, H. W. Lach, S. J. Birge, and J. P. Miller*
571 Using Physician Correspondence and Postcard Reminders to Promote Mammography Use *S. H. Taplin, C. Anderman, L. Grothaus, S. Curry, and D. Montano*
575 The Characteristics of Northern Black Churches with Community Health Outreach Programs *S. B. Thomas, S. C. Quinn, A. Billingsley, and C. Caldwell*

CHILD VIOLENCE AND INJURY

580 The Impact of the Safe Kids/Healthy Neighborhoods Injury Prevention Program in Harlem, 1988 through 1991 *L. L. Davidson, M. S. Durkin, L. Kuhn, P. O'Connor, B. Barlow, and M. C. Heagarty*
587 Low-Income Neighborhoods and the Risk of Severe Pediatric Injury: A Small-Area Analysis in Northern Manhattan *M. S. Durkin, L. L. Davidson, L. Kuhn, P. O'Connor, and B. Barlow*
593 Child and Adolescent Drownings in Harris County, Texas, 1983 through 1990 *C. L. Warneke and S. P. Cooper*
599 The Effects of Recall on Estimating Annual Nonfatal Injury Rates for Children and Adolescents *Y. Harel, M. D. Overpeck, D. H. Jones, P. C. Scheidt, P. E. Bijur, A. C. Trumble, and J. Anderson*
606 Characteristics of Adolescent Work Injuries Reported to the Minnesota Department of Labor and Industry *D. L. Parker, W. R. Carl, L. R. French, and F. B. Martin*
612 Factors Associated with the Use of Violence among Urban Black Adolescents *R. H. DuRant, C. Cadenhead, R. A. Pendergrast, G. Slavens, and C. W. Linder*

(continued on page 529)

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

April 1994, Vol. 84, No. 4 ISSN: 0090-0036 Established 1911

618 Aggression and Fighting Behavior among African-American Adolescents: Individual and Family Factors *N. U. Cotten, J. Resnick, D. C. Browne, S. L. Martin, D. R. McCarraher, and J. Woods*

623 Traumatic Child Death and Documented Maltreatment History, Los Angeles *S. B. Sorenson and J. G. Peterson*

628 Factors That Influence Clinicians' Assessment and Management of Family Violence *V. P. Tilden, T. A. Schmidt, B. J. Limandri, G. T. Chiodo, M. J. Garland, and P. A. Loveless*

634 Suicidal Ideation and Attempts among Chemically Dependent Adolescents *E. Y. Deykin and S. L. Buka*

Public Health Briefs

VARIOUS TOPICS

640 The Sexual Assault of Women at Work in Washington State, 1980 to 1989 *B. H. Alexander, G. M. Franklin, and M. E. Wolf*

643 Preventing Occupational Illness and Injury: Nurse Practitioners as Primary Care Providers *J. Lipscomb, B. Burgel, L. W. McGill, and P. Blanc*

CHILD VIOLENCE AND INJURY

646 Occupational Injury Deaths of 16- and 17-Year-Olds in the United States *D. N. Castillo, D. D. Landen, and L. A. Layne*

649 Head Injuries Incurred by Children and Young Adults during Informal Recreation *S. P. Baker, C. Fowler, G. Li, M. Warner, and A. L. Dannenberg*

653 A Cost-Benefit Analysis of Legislation for Bicycle Safety Helmets in Israel *G. M. Ginsberg and D. S. Silverberg*

657 Adolescent Occupational Injuries Requiring Hospital Emergency Department Treatment: A Nationally Representative Sample *L. A. Layne, D. N. Castillo, N. Stout, and P. Cutlip*

661 Risk Factors for Violent Behavior in Elementary School Boys: Have You Hugged Your Child Today? *J. L. Sheline, B. J. Skipper, and W. E. Broadhead*

Public Health Then and Now

664 Reflections on a Half Century of Injury Control *J. A. Waller*

Notes from the Field

671 Translating State Data into Local Programs: Targeted Research for Intervention Planning (TRIP) *S. L. Schensul, M. R. Eisenberg, J. Glasgow, and J. Huettner*

672 The Rhode Island Sun Smart Project: Skin Cancer Prevention Reaches the Beaches *J. S. Rossi, L. M. Blais, M. A. Weinstock*

Other Journal Departments

530 APHA Masthead
535 Advertisers' Index
536 What *AJPH* Authors Should Know
537 Journal Masthead
675 Letters to the Editor
678 Book Corner
682 Association News: Leadership Roster
696 Job Opportunities

American Public Health Association

American Journal of Public Health

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 15th St, NW, Washington, DC 20005; (202) 789-5600; TDD (202) 789-5673. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*
Helen Rodriguez-Trias, MD, *Immediate Past President*
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Leonard M. Schuman, MD, *Vice-President (USA)*
John E. F. Hastings, MD, MPH, *Vice-President (Canada)*
Oscar Feo, MD, MSc, *Vice-President (Latin America)*
Jay H. Glasser, PhD, *Treasurer*
Allan G. Rosenfield, MD, *Chair, Executive Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*
Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*
Carol Easley Allan, PhD, RN (1997)
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Patricia P. Evans, MPH (1994)
Eugene Feingold, PhD, JD, *President*
Jay H. Glasser, PhD, *Treasurer*
Alan R. Hinman, MD, MPH (1995)
Margo D. Kerrigan, MPH (1995)
Joanne Lukomnik, MD (1996), *Vice-Chair*
Steven Uranga McKane, DMD, MPH (1995)
Carlos W. Molina, EdD (1997)
Linda Rae Murray, MD, MPH (1996)
Hila Richardson, DrPH, MPH (1996)
Helen Rodriguez-Trias, MD, *Immediate Past President*
Reed V. Tuckson, MD (1994)
Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*
Frances Kissling, *Chair, Action Board*
George E. Pickett, MD, MPH, *Speaker, Program Development Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*

The ***American Journal of Public Health*** (*AJPH*), published monthly, is the official Journal of the American Public Health Association, Inc, 1015 15th St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005. *AJPH* allotted $40; *The Nation's Health* allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters. Printed in the United States.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on the first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, *AJPH* Reprints, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the Journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The ***American Journal of Public Health*** is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

Medicaid Physician Payment Reform: Using the Medicare Fee Schedule for Medicaid Payments

Anne Lenhard Reisinger, PhD, David C. Colby, PhD, and Anne Schwartz, MHS

ABSTRACT

Objectives. The purpose of this article is to provide estimates of the costs of basing Medicaid physician payment levels on the new resource-based Medicare Fee Schedule. Two possible policy options are considered: setting all Medicaid physician fees at the Medicare Fee Schedule level and setting only office visit fees at the new Medicare levels.

Methods. Data on Medicaid physician fees, use patterns, and the Medicare Fee Schedule are used to develop state-level estimates of expenditure changes under each option.

Results. Setting Medicaid rates at the Medicare Fee Schedule level could increase expenditures by $3.2 to $4.1 billion nationally; the other option would result in substantially lower increases in expenditures. Because of the current variations in Medicaid physician fees and in the breadth of eligibility across states, the cost of adopting the Medicare Fee Schedule varies considerably among states.

Conclusions. Adopting the new Medicare Fee Schedule for Medicaid payments, proposed by policymakers as a way to increase access to appropriate medical care, could double physician expenditures in some states. Adoption of more limited versions of the fee schedule might achieve some access gains at lower costs. (*Am J Public Health.* 1994;84:553–560)

Introduction

The Clinton administration's commitment to expand access to medical care and contain the growth in medical care costs has accelerated the debate over health care financing reform. Reform will no doubt include changes in the structure of the Medicaid program, although the scope of these changes is uncertain. While national reform could eliminate Medicaid as it currently exists, it more likely will change the program incrementally, gradually eliminating the differences between Medicaid and other insurance plans. This article discusses the cost of one such incremental change, the implementation of a resource-based physician fee schedule like that used in the Medicare program.

Medicaid was enacted in 1965 to improve the existing Kerr-Mills medical assistance program, expanding eligibility beyond the elderly to include other low-income Americans and increasing the federal share of program financing. Over the past 28 years, however, the program has grown beyond its original purpose. Today, Medicaid is viewed by many as a mechanism for increasing access to mainstream medical care for low-income Americans.

Although barriers to mainstream care remain, the Medicaid program has been largely successful in increasing access to health care services and improving the health status of those enrolled in the program.[1] After other differences have been controlled, the eligible poor receive more physician visits, more admissions to the hospital, and more hospital days than poor persons not covered by Medicaid.[2] When beneficiaries' eligibility for Medicaid terminates or when there are program cuts, access to health care diminishes and health status deteriorates.[3,4]

Despite these successes, there are problems with the Medicaid program. Expenditures for the program are rising rapidly and are expected to double between 1990 and 1995. Only about half of the poor are eligible for Medicaid, and the extent of coverage varies by state.[5,6] Moreover, even the enrolled poor are not guaranteed access to medical care; beneficiaries often find it difficult to locate a physician who will provide services, because many Medicaid fees are low.

Recently Congress passed legislation that ameliorates some of these problems. Medicaid eligibility has been expanded. For example, states are required to share Medicare costs for Medicare beneficiaries whose incomes are below the poverty line. States are required to provide Medicaid eligibility to pregnant women and children under 6 years of age in families with incomes at or below 133% of the poverty level and to all other children who live in families with incomes at or below the poverty line. With regard to physician payment, the Omnibus Reconciliation Act of 1989 required that states provide documentation that payment rates for obstetric and pediatric services were sufficient to ensure access comparable to that of the general population.

Policymakers are now considering a number of strategies for solving other problems with the Medicaid program, including changes in Medicaid physician payments. The available empirical evidence suggests that raising physician fees may encourage the provision of office-based physician services in lieu of hospital outpatient department and emergency room care. If changes in fee levels encourage physicians to treat more Medicaid patients in more appropriate settings, physician payment policy

Anne Lenhard Reisinger is with the School of International and Public Affairs, Columbia University, New York, NY. David C. Colby and Anne Schwartz are with the Physician Payment Review Commission, Washington, DC.

Requests for reprints should be sent to Anne Lenhard Reisinger, PhD, School of International and Public Affairs, Columbia University, 420 W 118th St, Room 1425, New York, NY 10027.

Note. The views expressed are the authors', and no endorsement by the Physician Payment Review Commission is intended or should be inferred.

may simultaneously address but will not entirely solve the problem of access for Medicaid beneficiaries.

The enactment of the resource-based Medicare Fee Schedule has created an impetus for similar reforms in Medicaid. Adoption of the Medicare Fee Schedule for Medicaid physician payment would have the advantages of increasing the overall level of payment and of realigning relative payments for primary care services and procedural services.

This article discusses the costs of two options for Medicaid physician payment reform based on the Medicare Fee Schedule. These options, which establish various floors on Medicaid physician payments, are adopting the full Medicare Fee Schedule and applying the Medicare Fee Schedule only to specific services such as visits.

Medicaid Physician Payment and Access to Care

Both academic and policy research have demonstrated access problems for Medicaid beneficiaries, especially in states with low fees. A 1990 survey of state Medicaid directors indicated that low physician participation is a problem for many states. Forty-three states identified participation problems, and more than half (27) identified problems with both geographic and specialty distribution of participating physicians.[7] Other data sources confirm variation in Medicaid physician participation by geographic area and specialty.[8,9]

Although there are a number of reasons why physicians limit their Medicaid participation, 30 states reported that low physician fees are the primary reason for low participation. The high cost of malpractice insurance premiums and complex billing procedures were also thought to be important causes of low participation.[7]

Health service researchers have demonstrated that low physician fees discourage physician participation in the Medicaid program. Higher Medicaid fees increase both the likelihood that a physician will treat Medicaid patients and the number of patients a physician treats.[10–17]

The positive effect of higher Medicaid fees on physician participation suggests that Medicaid and other insurers compete for physicians through their fees. When competing insurers pay higher fees than Medicaid, physicians are less likely to treat Medicaid patients.[12–14,16–18] This does not mean, however, that Medicaid patients will go without care. Rather, they may find other physicians who are willing to treat them, such as those with high-volume Medicaid practices, or they may obtain their care in other settings such as hospital outpatient departments, emergency rooms, and local health department clinics.

The extent of physician participation in Medicaid is a proxy measure of beneficiary access to care. Research that has directly measured the effect of fee levels on beneficiaries' use of medical care suggests that while Medicaid physician fee levels are not related to the number of services Medicaid beneficiaries receive, they are predictive of the site of service. Beneficiaries in states with higher fees are more likely to receive services in physicians' offices, while beneficiaries in states with lower fees are more likely to receive services in emergency rooms or outpatient departments.[19–22] This prompts concern because care provided in emergency rooms or outpatient departments is more expensive than that provided in physicians' offices and does not promote continuity of care.[23,24] In addition, inappropriate use of emergency rooms results in overcrowding, increased waiting times, and restricted access to care in that setting.[25–27]

As Congressman Henry Waxman has noted, "merely having a Medicaid card does not, in and of itself, assure access to physician services."[28] Since Medicaid purchases physicians' services in the fee-for-service market, both the level of physician fees and administrative procedures influence whether physicians will choose to treat Medicaid beneficiaries and how many they will treat.

Physician Payment Options: Adopting the Medicare Fee Schedule

Raising Medicaid fees would improve access to care provided in physicians' offices and provide mainstream care at lower costs per unit of service. Setting Medicaid fees at Medicare Fee Schedule levels would involve two major changes in Medicaid payments to physicians. First, it would require raising most fees in most states. Second, Medicaid payments to physicians would be based on resource costs. Under the Medicare Fee Schedule, Medicare fees for evaluation and management services (visits and consultations) have increased; fees for surgical and other technical procedures have decreased. Relative Medicaid physician fees would probably experience similar shifts.

Beneficiary access will improve if physicians respond to increases in fees by increasing their Medicaid participation and providing more office-based services to Medicaid beneficiaries. The extent of such improvements may vary among states because of differences in current payment levels. Access gains could be larger in states with particularly low Medicaid fees.

Equity among physicians might also be improved by reducing the current economic disparity between physicians with substantial Medicaid caseloads relative to their peers. Administratively, this would relieve Medicaid programs from the burden of setting fees.

The costs of physician payment reform options will weigh heavily in the debate over reform. States, already burdened with escalating Medicaid expenditures, may be reluctant to adopt reforms that increase those costs even more. Since Medicaid expenditures for physician services constitute only about 6% of total Medicaid expenditures, the financial impact of these options is relatively small. The following section estimates the initial costs of two physician payment reforms that would link Medicaid payments to the Medicare Fee Schedule.

Estimating Expenditure Changes

Estimated changes in Medicaid expenditures were calculated separately for beneficiaries dually eligible for Medicaid and Medicare and for all other beneficiaries. The two populations were separated because states are responsible for the full payment for nondually eligible beneficiaries and may elect to pay the 20% Medicare copayment for dually eligible beneficiaries. Twenty-four states pay the full Medicare coinsurance for dually eligible beneficiaries; the other states and the District of Columbia pay Medicare coinsurance for fees up to state Medicaid fee levels.

Nondually eligible beneficiaries. Data from Health Care Financing Administration (HCFA) Form 2082 were used to estimate expenditures for this population. Since expenditures on Form 2082

were for 1991, they were inflated to reflect the actual expenditures in 1992 (as recorded on HCFA Form 64).

The price difference between Medicaid and the Medicare Fee Schedule was measured by an index of the ratio of Medicare Fee Schedule to Medicaid fees for the same physician services averaged across nine service families. (Because Medicaid physician fees for only 16 services were used, fees were grouped into families to allow generalization from several fees to larger groups of services. The estimations therefore assume that these few fees represent a much larger set of services. The service families are office visits, hospital visits, emergency room visits, consultations, psychiatric services, electrocardiograms, x-ray services, obstetrical services, and surgical services and other procedures. Average Medicaid and Medicare Fee Schedule fees for each family of services were estimated on the basis of one or several physician services. The service families included the following Current Procedural Terminology (CPT) codes: limited [90050], intermediate [90060], and comprehensive [90020] office visits; initial [90220] and subsequent [90260] hospital visits, emergency room visits [90515]; initial consultation [90620]; individual psychotherapy [90844]; electrocardiogram [93000]; chest x-rays [71020, global fee and professional component]; vaginal delivery [59410] and delivery by cesarean section [59500]; cataract removal [66984]; hysterectomy [58150]; and upper gastrointestinal endoscopy [43235].) Medicaid fees for state fiscal year 1989 were drawn from the Physician Payment Review Commission survey of state Medicaid programs. Estimated Medicare Fee Schedule fees for those services were based on a simulation of Medicare Fee Schedule payments for 1988. The conversion factor for the Medicare Fee Schedule is assumed to be budget neutral to Medicare allowed charges for 1988 minus 6.5%. The reduction of 6.5% reflects the HCFA decision to include an adjustment for volume responses in its implementation of the Medicare Fee Schedule.[29]

The index for each individual state is a weighted mean of the fee ratios. The weights were determined by the percentage of Medicaid physician expenditures attributable to all possible service codes in each family of services. These percentages were calculated from the HCFA Medicaid Statistical Information System for 1989.

TABLE 1—Setting Medicaid Physician Fees at Medicare Fee Schedule (MFS) Levels, 1992[a]: Estimated Changes in Total Medicaid Physician Expenditures, by State

	Physician Expenditures 1992, $ Millions	Estimated Expenditures under MFS Levels, 1992, $ Millions	Change in Physician Expenditures, %	Change in Total Medicaid Expenditures, %
Alabama	91.56	117.34	28	2
Alaska	36.34	36.25	0	0
Arkansas	99.38	99.13	0	0
California	902.12	1241.06	38	3
Colorado	67.43	106.74	58	4
Connecticut	45.55	63.85	40	1
Delaware	8.91	13.42	51	2
District of Columbia	22.99	37.30	62	2
Florida	408.91	489.98	20	2
Georgia	317.32	315.75	0	0
Hawaii	33.50	35.10	5	1
Idaho	25.34	31.97	26	2
Illinois	264.09	483.57	83	5
Indiana	144.74	144.52	0	0
Iowa	79.75	83.89	5	0
Kansas	47.91	68.26	42	3
Kentucky	246.75	472.42	91	12
Louisiana	203.67	328.38	61	4
Maine	30.17	54.90	82	3
Maryland	244.49	445.94	82	11
Massachusetts	169.99	185.05	9	0
Michigan	249.82	423.34	69	5
Minnesota	93.59	99.81	7	0
Mississippi	115.29	210.68	83	9
Missouri	54.56	102.73	88	2
Montana	25.37	30.78	21	2
Nebraska	41.77	41.69	0	0
Nevada	43.92	43.92	0	0
New Hampshire	10.36	18.53	79	1
New Jersey	73.16	214.78	194	3
New Mexico	62.73	75.11	20	2
New York	224.61	798.33	255	3
North Carolina	211.32	243.57	15	1
North Dakota	13.79	18.21	32	2
Ohio	266.34	393.51	48	3
Oklahoma	80.91	91.65	13	1
Oregon	63.53	80.35	26	2
Pennsylvania	161.88	312.95	93	3
Rhode Island	9.32	18.41	98	1
South Carolina	106.47	138.83	30	2
South Dakota	17.39	22.60	30	2
Tennessee	244.33	278.71	14	1
Texas	525.50	598.52	14	1
Utah	39.15	49.82	27	3
Vermont	15.45	21.05	36	2
Virginia	169.35	241.96	43	5
Washington	168.72	249.05	48	4
West Virginia	86.32	210.92	144	13
Wisconsin	39.53	49.90	26	1
Total	6705.41	9934.54	48	3

Source. Physician expenditures by state are estimated from the 1991 Health Care Financing Administration (HCFA) Form 2082 inflated to 1992 amounts based on HCFA Form 64.
[a]Assumes no change in the volume of services provided by physicians in response to fee changes.

This fee index was then multiplied by each state's Medicaid physician expenditures for the nondually eligible population in 1989 (the amount Medicaid would have spent if it paid Medicare Fee Schedule prices). The model estimates

TABLE 2—Setting Medicaid Fees at Medicare Levels, 1992[a]: Estimated Changes in State and Federal Medicaid Physician Expenditures, by State

	Estimated Expenditure Change, $ Millions		
	Federal	State	Total
Alabama	18.79	6.98	25.77
Alaska	−0.04	−0.04	−0.09
Arkansas	−0.19	−0.06	−0.25
California	169.48	169.48	338.96
Colorado	21.53	17.77	39.30
Connecticut	9.15	9.15	18.31
Delaware	2.26	2.26	4.51
District of Columbia	7.16	7.16	14.31
Florida	44.34	36.73	81.07
Georgia	−0.97	−0.60	−1.57
Hawaii	0.80	0.80	1.60
Idaho	4.85	1.77	6.62
Illinois	109.74	109.74	219.48
Indiana	−0.14	−0.08	−0.23
Iowa	2.70	1.45	4.15
Kansas	12.05	8.30	20.35
Kentucky	164.34	61.34	225.68
Louisiana	94.08	30.63	124.71
Maine	15.43	9.30	24.73
Maryland	100.73	100.73	201.45
Massachusetts	7.53	7.53	15.06
Michigan	96.15	77.38	173.53
Minnesota	3.39	2.84	6.22
Mississippi	76.31	19.09	95.40
Missouri	29.30	18.86	48.16
Montana	3.88	1.53	5.42
Nebraska	−0.05	−0.03	−0.08
Nevada	0	0	0
New Hampshire	4.08	4.08	8.16
New Jersey	70.81	70.81	141.62
New Mexico	9.20	3.18	12.38
New York	286.86	286.86	573.71
North Carolina	21.45	10.80	32.25
North Dakota	3.22	1.20	4.42
Ohio	77.10	50.07	127.17
Oklahoma	7.60	3.14	10.74
Oregon	10.68	6.13	16.81
Pennsylvania	85.87	65.20	151.07
Rhode Island	4.84	4.25	9.09
South Carolina	23.51	8.85	32.36
South Dakota	3.78	1.43	5.20
Tennessee	23.52	10.86	34.38
Texas	46.86	26.15	73.02
Utah	8.01	2.65	10.66
Vermont	3.43	2.16	5.59
Virginia	36.31	36.31	72.61
Washington	44.16	36.16	80.33
West Virginia	96.78	27.81	124.59
Wisconsin	6.26	4.11	10.37
Total	1866.94	1362.19	3229.13

[a]Assumes no change in the volume of services provided by physicians in response to fee changes.

the first order effect of adopting the Medicare Fee Schedule, assuming that increased fees have no effect on the volume of services provided.

The estimates assume that states with average Medicaid fees above those of the Medicare Fee Schedule will not decrease their fees to Medicare Fee Schedule levels. If they do so, the total change will be somewhat lower than estimated. Changes in office visit expenditures were estimated by including only office visit fees in the index of Medicare Fee Schedule to Medicaid fees and applying this ratio to the expenditures attributable to office visits. (The estimates assume that approximately 30% of expenditures are attributable to office visits. This estimate is based on data from the Medicaid Statistical Information System for 1989.)

Dually eligible beneficiaries. Projected total Medicare Fee Schedule expenditures for dually eligible beneficiaries were calculated in two ways. For states that pay the full coinsurance, the percentage change in Medicare expenditures under the Medicare Fee Schedule relative to Medicare allowed charges was simulated and multiplied by 1992 expenditures for the dually eligible beneficiaries. Second, for states that do not pay the full Medicare coinsurance, the fee index used for the nondually eligible beneficiaries was multiplied by current expenditures.

The estimates of changes in expenditures assume that all states would be required to pay the 20% copayment for dual eligible beneficiaries so that services for all Medicaid beneficiaries are paid at the Medicare Fee Schedule rate. Since state Medicaid programs' policies for this population follow Medicare's, both increases and decreases in payments are included in the expenditure estimates for this population.

Setting Medicaid Fees at Medicare Fee Schedule Levels

Setting Medicaid fees at Medicare Fee Schedule levels would increase total Medicaid physician service expenditures for current service volumes from $6.71 billion to $9.93 billion (Table 1). This additional $3.23 billion would increase total Medicaid expenditures (all services including physicians, hospitals, and long-term care) by about 3%. Under the current matching formula, the state share would be approximately $1.36 billion, and the federal share would be an estimated $1.87 billion.

The cost of raising Medicaid physician fees to Medicare Fee Schedule levels would vary among states because of current differences in Medicaid fee levels and use across states. Assuming no change in the volume of services, Medicaid physician expenditures in Alaska, Arkansas, Georgia, Indiana, Nebraska, and Nevada would experience little change, while expenditures in New York, New Jersey, and West Virginia would increase dramatically. Under this op-

tion, about 18% of the total growth in Medicaid physician expenditures would occur in New York.

The state shares of these increased expenditures would also vary because of differences in the current Medicaid matching formula (Table 2). For example, while Illinois and Mississippi would experience comparable percentage increases in expenditures (83%), Illinois would pay about 50% of the cost of its growth in physician expenditures and Mississippi would pay only about 20%.

Raising Medicaid fees to Medicare levels may be difficult to achieve in the short term. Even at current fee levels, states are already straining under the weight of growing Medicaid expenditures. In view of these financial pressures, policymakers may wish to consider more limited options for reforming Medicaid physician payment.

Increasing Fees for Selected Services

A second option is to raise fees for only certain types of services, such as office visits. A number of states have already implemented fee increases for primary care services and deliveries. This option targets increases in expenditures to those services for which access is most problematic.

Raising Medicaid fees for office visits is a particularly appealing policy option. Previous research has shown that low physician fees lead to increased care in hospital outpatient departments and emergency rooms. Raising office visit fees could encourage the substitution of office care for outpatient department and emergency room care in appropriate instances. Raising Medicaid office visit fees to Medicare Fee Schedule levels would increase physician expenditures by approximately $1.12 billion, or less than 1% of total Medicaid expenditures (see Tables 3 and 4). The federal share of this increase would be about $661 million; state expenditures would rise by approximately $460 million.

Again, the expenditure increases in individual states vary because of differences in current Medicaid fee levels, in the numbers of beneficiaries in each state, and in the use of office-based physician services. More than 25% of the total increase in expenditures due to this option would occur in California and New York. Expenditures for physician services in New York and West Virginia would increase by considerably more than 50%. Alaska, Arkansas, Massachusetts, and Nevada, which already have Medicaid office visit fees very near or above Medicare Fee Schedule levels, would incur negligible expenditure changes

TABLE 3—Setting Medicaid Office Visit Fees at Medicare Fee Schedule (MFS) Levels, 1992[a]: Estimated Changes in Total Medicaid Physician Expenditures, by State

	Physician Expenditures, 1992, $ Millions	Estimated Expenditures under MFS Levels, 1992, $ Millions	Change in Physician Expenditures, %	Change in Total Medicaid Expenditures, %
Alabama	91.56	103.19	13	1
Alaska	36.34	36.31	0	0
Arkansas	99.38	99.30	0	0
California	902.12	1056.99	17	1
Colorado	67.43	72.02	7	0
Connecticut	45.55	52.19	15	0
Delaware	8.91	10.69	20	1
District of Columbia	27.99	26.62	16	1
Florida	408.91	431.64	6	1
Georgia	317.32	330.67	4	1
Hawaii	33.50	35.02	5	0
Idaho	25.34	28.02	11	1
Illinois	264.09	361.15	37	2
Indiana	144.74	154.77	7	0
Iowa	79.75	88.51	11	1
Kansas	47.91	51.17	7	0
Kentucky	246.75	301.86	22	3
Louisiana	203.67	251.31	23	2
Maine	30.17	35.88	19	1
Maryland	244.49	277.77	14	2
Massachusetts	169.99	169.99	0	0
Michigan	249.82	303.97	22	1
Minnesota	93.59	103.48	11	1
Mississippi	115.29	144.94	26	3
Missouri	54.56	66.77	22	1
Montana	25.37	27.80	10	1
Nebraska	41.77	45.17	8	1
Nevada	43.92	43.92	0	0
New Hampshire	10.36	11.59	12	0
New Jersey	73.16	103.70	42	1
New Mexico	62.73	69.63	11	1
New York	224.61	363.71	62	1
North Carolina	211.32	225.59	7	1
North Dakota	13.79	17.73	29	2
Ohio	266.34	315.08	18	1
Oklahoma	80.91	87.62	8	1
Oregon	63.53	73.03	15	1
Pennsylvania	161.88	220.22	36	1
Rhode Island	9.32	10.81	16	0
South Carolina	106.47	115.38	8	1
South Dakota	17.39	18.96	9	1
Tennessee	244.33	256.61	5	1
Texas	525.50	593.36	15	1
Utah	39.15	43.35	11	1
Vermont	15.45	17.16	11	1
Virginia	169.35	193.90	14	2
Washington	168.72	183.40	9	1
West Virginia	86.32	148.38	72	7
Wisconsin	39.53	46.23	17	0
Total	6705.41	7826.56	17	1

Source. Physician expenditures by state are estimated from the 1991 Health Care Financing Administration (HCFA) Form 2082 inflated to 1992 amounts based on HCFA Form 64.
[a]Assumes no change in the volume of services provided by physicians in response to fee changes.

Although this option is more limited, several states, such as Indiana, would have greater increases in expenditures under this

TABLE 4—Setting Medicaid Office Visit Fees at Medicare Fee Schedule Levels, 1992[a]: Estimated Changes in State and Federal Medicaid Physician Expenditures, by State

	Estimated Expenditure Change, $ Millions		
	Federal	State	Total
Alabama	8.48	3.15	11.62
Alaska	−0.01	−0.01	−0.03
Arkansas	−0.57	−0.18	−0.74
California	77.43	77.43	154.87
Colorado	2.51	2.07	4.59
Connecticut	3.32	3.32	6.64
Delaware	0.89	0.89	1.78
District of Columbia	1.82	1.82	3.63
Florida	12.43	10.30	22.74
Georgia	8.25	5.10	13.35
Hawaii	0.76	0.76	1.52
Idaho	1.96	0.72	2.68
Illinois	48.53	48.53	97.07
Indiana	6.41	3.63	10.03
Iowa	5.70	3.06	8.76
Kansas	1.93	1.33	3.26
Kentucky	40.13	14.98	55.11
Louisiana	35.94	11.70	47.64
Maine	3.57	2.15	5.71
Maryland	16.64	16.64	33.28
Massachusetts	0.00	0.00	0.00
Michigan	30.01	24.15	54.15
Minnesota	5.38	4.51	9.89
Mississippi	23.72	5.93	29.65
Missouri	7.43	4.78	12.21
Montana	1.74	0.69	2.43
Nebraska	2.19	1.21	3.40
Nevada	0.00	0.00	0.00
New Hampshire	0.61	0.61	1.23
New Jersey	15.27	15.27	30.54
New Mexico	5.13	1.77	6.89
New York	69.55	69.55	139.10
North Carolina	9.49	4.78	14.27
North Dakota	2.86	1.07	3.94
Ohio	29.55	19.19	48.73
Oklahoma	4.75	1.96	6.71
Oregon	6.04	3.47	9.50
Pennsylvania	33.16	25.18	58.35
Rhode Island	0.79	0.69	1.49
South Carolina	6.48	2.44	8.91
South Dakota	1.13	0.43	1.56
Tennessee	8.40	3.88	12.28
Texas	43.55	24.31	67.86
Utah	3.15	1.04	4.20
Vermont	1.05	0.66	1.71
Virginia	12.28	12.28	24.56
Washington	8.07	6.61	14.48
West Virginia	48.20	13.85	62.05
Wisconsin	4.04	2.65	6.70
Total	660.65	460.50	1121.15

[a]Assumes no change in the volume of services provided by physicians in response to fee changes.

proposal than if they raised all fees to the Medicare level. The reason is that their fees for evaluation and management services are lower relative to Medicare fees than are their procedural and surgical fees.

Other more limited options are available. For example, setting Medicaid fees at a percentage of Medicare Fee Schedule levels (e.g., 80%) would have the advantage of shifting the relative values for all services and increasing fees in many states and would be less costly to implement. The potential drawback to this option, however, is that the remaining gap between Medicaid fees and those of other payers may be substantial and could continue to discourage office-based physicians from providing services. Since primary care fees would increase relatively more than fees for other types of services, however, this policy might succeed in promoting greater willingness of primary care physicians to participate in Medicaid.

Impact of Changes in Volume and Site of Service

The previous sections have presented estimates of the first order effect of moving Medicaid fees to Medicare Fee Schedule levels. Total expenditures also will be influenced by changes in the volume of services provided, the site of primary care, and the mix of services.

The effect of changes in fee levels on total expenditures will be influenced by physicians' decisions regarding the volume of services provided to Medicaid beneficiaries. Recently, the Congressional Budget Office estimated that there would be a 50% increase in volume for Medicaid beneficiaries if Medicaid rates were raised to Medicare levels.[30] Work by Long and his colleagues suggests that a 10% increase in Medicaid physician fees for visits would increase the number of office visits by 3%.[19]

If the volume of office visits increases, the number of emergency room and outpatient hospital visits may decrease. Using the National Health Interview Survey for 1989, we estimated the proportion of emergency room or outpatient department visits for Medicaid beneficiaries that could be considered necessary. Following Sloan's research, visits for certain emergency conditions and on weekends were classified as necessary.[31] Slightly fewer than 25% of these visits appear to be necessary. Therefore, by improving access to office-based care and developing other policy changes such as primary care case management, there is the potential to save approximately 75% of Medicaid expenditures for outpatient and emergency room visits.

Expenditure changes due to increased volume would be partially offset by decreases in the use of emergency rooms (Table 5). If, as the Congressional Budget Office estimates, there would be a 50% increase in the volume of services

TABLE 5—Changes in Medicaid Physician Expenditures under Different Assumptions, 1992

Option	No Volume Response			With Volume and Site of Service Changes		
	Total Expenditure Change, $ Billions	Change in Physician Expenditures, %	Change in Total Expenditures, %	Total Expenditure Change, $ Billions	Change in Physician Expenditures, %	Change in Total Expenditures, %
All fees at Medicare levels	3.2	48	3	4.1	61	4
Visit fees at Medicare levels	1.1	17	1	−0.6	−10	−1

and potential savings in emergency room expenditures are realized, total Medicaid physician expenditures for the policy option of raising all Medicaid fees to Medicare levels would be $4.1 billion (or a 4% increase in total Medicaid expenditures). Under the option of raising physician visit fees only, there would be an overall decrease in expenditures. While it is doubtful that all expenditures from unnecessary emergency room care could be eliminated, there would be additional savings not included in these estimates because beneficiaries are less likely to use inpatient care when the original site of care is a physician's office. Medicaid expenditures for inpatient hospital services would probably decrease with the shift in the site of visits.[25,26]

Discussion

The available empirical evidence suggests that Medicaid physician payment policy affects the medical care that beneficiaries receive, often creating incentives that drive Medicaid patients away from physicians' offices to hospital outpatient departments or emergency rooms. Changing the payment levels for physicians' services could alter these incentives and make physician office care more available to Medicaid beneficiaries. The path that Medicare policy has followed—paying physicians for services based on a resource-based relative value scale—has potential usefulness for the Medicaid program. Medicaid payments based on the Medicare Fee Schedule could realign the relative values of Medicaid physician payments and raise these payments.

These reforms could be costly to institute, however. Although there would be a small increase in total national Medicaid expenditures, the increased costs to states would vary tremendously because of current differences in Medicaid programs among the states. Physician expenditures in a few states could double.

Strategies such as setting Medicaid fees at a percentage of Medicare Fee Schedule payments or adopting the fee schedule for certain types of services could be used to increase access to care at lower cost. These more limited approaches could be used to target increased expenditures to services for which the access problems are the most severe.

These options alone, however, may not improve access where it is most needed. In some locations, such as inner-city and rural areas, the present supply of physicians is too small to provide needed services. While it might be possible to raise fee levels enough to make some physicians willing to locate in underserved areas and to treat Medicaid patients, other factors may limit the magnitude of this response.[31] These factors include cultural and language differences, characteristics of these areas, and physicians' perceptions that Medicaid patients tend to have more serious medical problems, miss more appointments, and are less compliant than other patients. Some physicians are also wary that being labeled as a "Medicaid physician" may prompt other patients to seek care elsewhere. □

Acknowledgments

We wish to thank Lauren LeRoy and Paul Ginsburg for comments on the paper and Jansen Davis and Arlene Turner for computer programming.

References

1. Davis K, Schoen C. *Health and the War on Poverty.* Washington, DC: Brookings Institution; 1978.
2. Davis K, Reynolds R. The impact of Medicare and Medicaid on access to medical care. In Rosett RN, ed. *The Role of Health Insurance in the Health Services Sector.* New York, NY: National Bureau of Economic Research; 1976.
3. Lurie N, Ward NB, Shapiro MF, Brook RF. Termination from Medi-Cal—does it affect health? *N Engl J Med.* 1984;311:480–484.
4. Lurie N, Ward NB, Shapiro MF, Gallego C, Vaghaiwalla R, Brook RH. Termination of Medi-Cal benefits: a follow-up study one year later. *N Engl J Med.* 1986;314:1266–1268.
5. *Overview of Entitlement Programs.* Washington, DC: US Ways and Means Committee; 1991.
6. Holahan J, Zedlewski S, Winterbottom C. *Estimating the Cost of Medicaid Eligibility and Benefit Uniformity.* Washington, DC: The Urban Institute; 1990. Working Paper 3836-04.
7. *Physician Payment under Medicaid.* Washington, DC: Physician Payment Review Commission; 1991.
8. American Medical Association. Medicaid participation declines. *SMS Rep.* 1989;3:1–3.
9. Mitchell JB. Physician participation in Medicaid revisited. *Med Care.* 1991;29:645–653.
10. Davidson SM. Physician participation in Medicaid: background and issues. *J Health Polit Policy Law.* 1982;6:703–717.
11. Davidson SM, Perloff JD, Kletke P, Schiff DW, Connelly JP. Full and limited Medicaid participation among pediatricians. *Pediatrics.* 1983;72:552–559.
12. Hadley J. *Physician Participation in Medicaid.* Washington, DC: The Urban Institute; 1978. Working Paper 9-998-9.
13. Held P, Holahan J, Carlson C. *The Effect of Medicaid and Private Fees on Physician Participation in California's Medicaid Program, 1975–1978.* Washington, DC: The Urban Institute; 1983. Working Paper 1306-02-01.
14. Mitchell J. Medicaid participation by medical and surgical specialists. *Med Care.* 1983;21:929–938.
15. Perloff JD, Kletke PR, Neckerman KM. Recent trends in pediatrician participation in Medicaid. *Med Care.* 1986;24:749–760.
16. Sloan FA, Cromwell J, Mitchell JB. *Private Physicians and Public Programs.* Lexington, Mass: Lexington Books; 1978.
17. Yudkowsky BK, Cartland JDC, Flint SS. Pediatrician participation in Medicaid: 1978 to 1989. *Pediatrics.* 1990;85:567–577.
18. Mitchell J, Schurman R. Access to private obstetrics/gynecology services under Medicaid. *Med Care.* 1984;22:1026–1037.
19. Long SH, Settle RF, Stuart BC. Reim-

bursement and access to physicians' services under Medicaid. *J Health Economics.* 1986;5:235–251.
20. Cohen J. Medicaid policy and the substitution of hospital outpatient care for physician care. *Health Services Res.* 1989;24:33–66.
21. Rosenbach M. The impact of Medicaid on physician use by low-income children. *Am J Public Health.* 1989;79:1220–1225.
22. Cohen J. *Medicaid Physician Fees and Use of Physician and Hospital Services.* Chicago, Ill: University of Chicago; 1991. Dissertation.
23. Gold M. The demand for hospital outpatient services. *Health Services Res.* 1984;19: 383–411.
24. McDevitt R, Dutton B. *Expenditures for Ambulatory Episodes of Care: The Michigan Medicaid Experience.* Baltimore, Md: Health Care Financing Administration, Office of Research; 1987.
25. Grumbach K, Keane D, Bindman A. Primary care and public emergency department overcrowding. *Am J Public Health.* 1993;83:372–378.
26. Baker DW, Stevens CD, Brook RH. Patients who leave a public hospital emergency department without being seen by a physician. *JAMA.* 1991;266:1085–1090.
27. Bindman AB, Grumbach K, Keane D, Rauch L, Luce JM. Consequences of queuing for care at a public hospital emergency department. *JAMA.* 1991;266: 1091–1096.
28. Waxman H. Kids and Medicaid: progress but continuing problems. *Am J Public Health.* 1989;79:1217–1218.
29. Health Care Financing Administration. Medicare program; fee schedule for physicians' services. *Federal Register.* 1991; 56:59 502–59 819.
30. *Single-Payer and All-Payer Health Insurance Systems Using Medicare's Payment Rates.* Washington, DC: Congressional Budget Office; 1993.
31. Sloan F. The demand for physicians' services in alternative practice settings: a multiple logit approach. *Q Rev Economics Statistics.* 1978;18:41–62.
32. Fossett JW, Perloff JD, Peterson JA, Kletke PR. Medicaid in the inner city: the case of maternity care in Chicago. *Milbank Q.* 1990;68:111–141.

Submissions Invited for Public Health Policy Forum

The Journal department "Public Health Policy Form" is intended to present divergent views on important public health policy issues in a more extensive format than the Journal usually allows. Three kinds of material appear in the Forum: articles, not exceeding 4500 words; commentaries, not exceeding 2500 words; and editorials. Customarily, commentaries and editorials are solicited by the Journal's Contributing Editor or Editor. Articles for the Forum are selected from those submitted generally to the Journal or specifically to the Forum. In general, peer review is sought.

Those wishing to submit articles directly to the Public Health Policy Forum should send them to the Journal office as follows:

George A. Silver, MD, MPH
Contributing Editor, *AJPH* Public Health Policy Forum
1015 Fifteenth Street, NW
Washington, DC 20005

Please send five copies of the manuscript and follow the Journal's guidelines, "What *AJPH* Authors Should Know," printed in each issue of the Journal.

The scholarly merit and the scientific accuracy of all submissions are considered. Additional criteria are relevance to an important policy issue, timeliness, and clarity and coherence of the policy argument. Brevity also helps considerably.

Homicide, Handguns, and the Crime Gun Hypothesis: Firearms Used in Fatal Shootings of Law Enforcement Officers, 1980 to 1989

Garen J. Wintemute, MD, MPH

ABSTRACT

Objectives. Many policies seeking to limit handgun violence rest on the largely untested "crime gun hypothesis," which holds that subclasses of handguns differ in their risk for use in violent crime. This study tests that hypothesis for handguns used in homicides of law enforcement officers and describes the population of homicide-involved handguns.

Methods. A cross-sectional study was done of civilian (criminal) handguns used in homicides of law enforcement officers from 1980 to 1989. Life tables were generated for each year's cohort of new handguns to estimate gun-years at risk, analogous to person-years, for rate and relative risk calculations.

Results. Four hundred thirty-five deaths involved 428 civilian handguns. Revolvers were at greater risk than pistols. For both, risk was lowest for .22-caliber handguns. Risk was greatest for .32-caliber pistols and .38-caliber revolvers. Forty-six percent of handguns had a barrel length of 3 in or less.

Conclusions. Subclasses of handguns differ substantially in their risk for use in fatal shootings of law enforcement officers. Such epidemiological data may be useful in formulating efforts to prevent these and similar instances of firearm violence. (*Am J Public Health.* 1994; 84:561–564)

Introduction

Regulatory efforts to prevent firearm violence often address specified classes of firearms. In the 1960s, the proliferation and criminal use of inexpensive, small-caliber handguns—Saturday Night Specials—led to severe restrictions on their importation under the Gun Control Act of 1968. In 1990, similar concerns regarding certain high-capacity semiautomatic weapons led to an administrative ban on their importation.

Such policies are grounded in a belief in the validity of the "crime gun hypothesis," which holds that all firearms are not at equal risk for use in a violent crime. This hypothesis has clearly been validated in the case of handguns generally, as compared with rifles and shotguns. While handguns constitute perhaps 40% of firearms produced in the United States in the past 2 decades, they account for 80% of all violent crimes involving firearms.[1]

The crime gun hypothesis has not been rigorously tested for the various classes of handguns because the necessary data have been lacking. While information on the characteristics of handguns released into circulation—as approximated by data on handguns manufactured or imported—is compiled only on a national basis, no nationally based data on handguns actually used in violent crimes have been available.

Previous explorations of this issue have yielded conflicting results. In the 1970s, half or more of all handguns confiscated by police agencies were of small caliber, had short barrels, or both.[2,3] Later critics noted that many of these "crime guns" had not actually been used in violent crime; some had no criminal associations at all.[4,5] In contrast, two surveys of convicted felons suggested that such criminals most often use high-quality, high-caliber handguns, particularly if they rely heavily on firearms in committing crimes.[6,7] Both studies were based entirely on self-report; new data on handguns actually confiscated from violent career felons at the time of their arrest show that both large- and small-caliber handguns are in common use by this group.[8] None of these studies estimated the actual or relative risk of different classes of handguns being used in crime.

The present study attempts to provide a more exact, albeit narrowly focused test of the crime gun hypothesis. Its subjects are the civilian handguns used in homicides of law enforcement officers in the United States from 1980 to 1989; this is the only instance in which detailed, nationally based data on firearms actually used in violent crime have been compiled. Such data are compatible with the nationally based information used to estimate handgun availability.

I estimate absolute and relative risks for involvement in these fatal shootings by handgun type (revolver or pistol) and caliber, the two characteristics for which availability data are collected, and I test two hypotheses: (1) that risk will be positively associated with handgun caliber; and (2) that risk will be higher for pistols than for

Garen J. Wintemute is with the Department of Community and International Health, University of California, Davis.

Requests for reprints should be sent to Garen J. Wintemute, MD, MPH, University of California, Davis, The Glenn Dairy Building, 1700 Alhambra Blvd, Suite 106, Sacramento, CA 95816.

This paper was accepted August 10, 1993.

Editor's Note. See related editorial by Kaufer Christoffel (p 539) and annotation by Kellermann (p 541) in this issue.

revolvers. As for the first hypothesis, small area studies have found an association between larger caliber and higher case-fatality rates for assaultive violence involving handgun use.[2,4,9] As for the second, pistols have been reported anecdotally to be used with increasing frequency in violent crime, including the crime at issue here. Pistols carry more ammunition than revolvers do, typically between 9 and 16 rounds, and automatically bring a new round into firing position each time a round is discharged. Revolvers generally carry only 5 to 6 rounds, and many must be recocked manually each time the gun is to be fired.

Finally, I describe this population of homicide-involved handguns with regard to characteristics for which the existing data do not permit an estimation of risk.

Methods

Data Sources

Data on all firearm homicides of law enforcement officers in the United States from 1980 to 1989 were provided by the Uniform Crime Reporting Program of the Federal Bureau of Investigation (FBI). Information in the dataset is collected from local law enforcement agencies, FBI field divisions, and the agency's Bureau of Justice Assistance. Reports are cross-checked for validity and completeness.[10]

Data on new handgun availability were provided by the Bureau of Alcohol, Tobacco and Firearms, which collects such information from manufacturers and importers by handgun type and caliber, but not by barrel length or ammunition capacity. Information on handguns manufactured in the United States was first collected in 1973; comparable data on imported handguns were not collected until 1988. Data on exported handguns, which account for about 10% of domestic manufactures, do not exist.

Data Management

Not all records were complete, presumably owing to failure to obtain the weapon involved. Missing information was sought from manufacturer catalogs, lay publications, and scholarly manuals of firearms identification.[11,12] Data were added for 52 handguns when this supplemental review established that only one value for the missing variable was possible. Residual frequencies for missing data were greater than 5% for gun type (7%), manufacturer (25%), and barrel length (34%).

Seven shootings were reported in which more than one law enforcement officer appeared to have been killed with the same handgun. Only one observation was retained.

Analysis

Numerators for rate and relative risk estimates were taken directly from the FBI data. Because imported handguns were missing from the denominator dataset until 1988, homicides involving such weapons were excluded until that year.

Denominators were expressed as gun-years at risk, an analog to person-years, for each gun type and caliber. Gun-years at risk accumulated from 1980 to 1989 were estimated for each annual cohort of new handguns produced from 1973 through 1989, specific to handgun type and caliber, by generating demographic life tables.[13,14]

The size of each annual cohort at inception was approximated by the number of handguns manufactured for 1973 to 1987, and by the number of handguns manufactured plus the number of handguns imported for 1988 to 1989. The annual attrition rate was also estimated, as nationally based data on the persistence of handguns in circulation are not available. Police handgun confiscation studies from the 1970s found that approximately half of these weapons had been manufactured within 3 to 5 years of confiscation.[2,3,15] The life tables were therefore first constructed under the assumption that the annual rate of attrition was 0.10 for each gun type and caliber. Rates were calculated under this model.

To estimate the sensitivity of the results to variations in the survival model, gun-years at risk for each annual cohort of new handguns were also calculated for rates of attrition of 0.05 and 0.20. Confidence intervals (CIs) for relative risk (RR) estimates were derived by the method of Katz[16] using Confidence Interval Analysis software.[17]

Results

There were 735 firearm homicides of law enforcement officers from 1980 to 1989, of which 435 (59%) were committed with the 428 civilian (criminal) handguns that are the subjects of the present study. Of these guns, 296 (69%) were revolvers, 104 (24%) were pistols, and 28 (7%) were of unknown type. Most handguns (53%) were of medium caliber (.38 or 9 mm). Small-caliber (.22 to .32) handguns accounted for 20% of all weapons.

Short-barreled handguns predominated among both pistols and revolvers; of 288 guns for which information was available, 134 (47%) had a barrel length of 3 in or less. Such weapons accounted for 78% of all .22- to .32-caliber handguns and 54% of all .38-caliber or 9-mm handguns.

Firearms manufacturers based in the United States accounted for 82% of the 322 handguns for which a manufacturer was identified; the four most frequently identified firms produced 59% of these weapons.

Revolvers were at greater risk than pistols (RR = 1.40; 95% CI = 1.10, 1.77). For pistols and revolvers alike, handgun caliber was associated with large differences in their rates of use in these events (Figure 1). Rates were lowest for .22-caliber handguns. Among pistols, .32-caliber weapons were at greatest risk (RR = 15.3; 95% CI = 5.56, 41.9), but these relatively uncommon weapons accounted for only 9% of all homicide-involved pistols. Nearly half (49%) of all pistols were of .38 or 9-mm caliber (RR = 7.28; 95% CI = 3.65, 14.5). Among revolvers, risk was highest for .38-caliber weapons (RR = 9.02; 95% CI = 5.31, 15.3). This group accounted for 57% of all revolvers of known caliber and 42% of all handguns for which both gun type and caliber were known. Rates for the largest caliber handguns were lower than those for weapons of intermediate caliber.

The relative risk estimates were minimally sensitive to fourfold variations in the hypothesized rate at which handguns disappear from circulation. The ranking of handgun types and calibers by their relative risk for use in these events was not affected (Table 1).

Discussion

The crime gun hypothesis appears to be valid for the specific crime of homicide of a law enforcement officer. Among commonly available handguns, differences in risk for use in these fatal shootings are associated with specific design characteristics. Some of these differences are substantial; all are only modestly affected by large variations in

the hypothesized rate at which handguns are removed from circulation.

The finding that small-caliber handguns are at relatively low risk was expected. Handguns used in homicides are substantially more likely than handguns used in nonfatal assaults to be of large caliber.[2,4,12] The finding of low risk at low caliber may be specific to homicide, reflecting the lower lethality of these handguns.

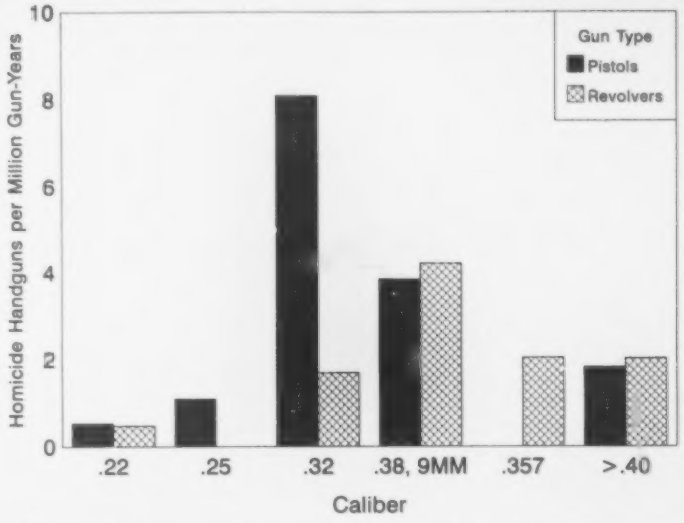
The relatively low rates among the highest caliber weapons were not expected but may be partly explained by price elasticity in handgun acquisition decisions by criminal users. High-caliber handguns are often substantially more expensive than low-caliber weapons—by as much as 50% or more at retail list prices and perhaps more at street pricing.

The predominance of handguns with short barrels—a key factor in weapon concealability—is consistent with earlier reports. Police studies have consistently found a high frequency of short-barreled weapons among confiscated handguns.[2,3,4] Convicted felons report that concealability was one of the chief factors influencing their handgun selection.[7] A Florida survey of convicted felons and gun owners in the general population found that the latter group was far less likely to report owning a handgun with a barrel length of 3 in or less (odds ratio = 0.22; 95% CI = 0.15, 0.34; this author's calculation).[6]

The quality of the data available for this study raises several caveats. The FBI data frequently omit handgun manufacturer and barrel length. However, there is no a priori reason to expect that values for these variables are related to the frequency of missing data, and no bias is expected. The Bureau of Alcohol, Tobacco and Firearms data describing handguns in the United States do not include characteristics such as barrel length; the association between these characteristics and risk for use in firearm violence cannot be quantified.

A direct measure of guns sold at retail does not exist. Over a brief period of time, fluctuations in the inventory of unsold handguns could result in a significant difference between the number of handguns available for sale and the number actually sold (or stolen) and entering the population at risk for criminal use. However, it is unlikely that such a difference would persist across the 17-year period used here.

Similarly, no nationally based data exist on the persistence of handguns in circulation. The robustness of the relative risk estimates across widely varying hypothesized rates of attrition suggests that a reasonable assessment of relative risk, if not of absolute risk, can be made without this information. However, the results depend on the design of the model used to generate them. We have assumed that all subclasses of handguns are lost from circulation at the same rate. This may not be the case. In particular, the guns most at risk for use in violent crime may also be most at risk for confiscation and for abandonment. If this is true, then gun-years for these weapons have been overestimated and the relative risks presented here are artificially low.



Note: Homicide handguns per million gun-years is defined as the number of guns used in homicides divided by the estimated total number of handguns times their duration in circulation (in millions of years). See the Methods section.

FIGURE 1—Rates of handgun use in homicides of law enforcement officers, by caliber and gun type, United States, 1980 to 1989.

TABLE 1—Relative Risk for the Involvement of Handguns in Homicides of Law Enforcement Officers, by Estimate of the Proportion of Handguns Lost from Circulation Each Year: United States, 1980 to 1989

	0.05		0.10		0.20	
Gun Type/ Caliber	Relative Risk	95% Confidence Interval	Relative Risk	95% Confidence Interval	Relative Risk	95% Confidence Interval
Revolver						
.22	1		1		1	
.32	3.58	1.70, 7.52	3.61	1.72, 7.58	3.63	1.73, 7.63
.38	9.07	5.34, 15.4	9.02	5.31, 15.3	8.92	5.25, 15.1
.357	4.48	2.56, 7.84	4.36	2.49, 7.62	4.17	2.39, 7.29
>.40	4.51	2.36, 8.65	4.31	2.25, 8.26	4.35	2.27, 8.32
Pistol						
.22	1		1		1	
.25	2.03	0.95, 4.33	2.05	0.96, 4.38	2.08	0.98, 4.45
.32	16.90	6.17, 46.5	15.30	5.56, 41.9	12.80	4.68, 35.2
.38/9 mm	7.36	3.69, 14.7	7.28	3.65, 14.5	7.15	3.58, 14.3
>.40	3.45	1.51, 7.87	3.45	1.51, 7.85	3.45	1.51, 7.86

Other limitations arise from the absence of more generally applicable information. First, handguns used in homicides of law enforcement officers may not be representative of all handguns used in homicides. While .22- to .32-caliber handguns made up 20% of the weapons in this series, they constituted 33% to 52% of homicide-involved handguns in four major cities in studies conducted during the 1970s.[2,9] The creation of a firearm fatality reporting system has recently been recommended[18] and could be implemented with relative ease.

Second, the handguns most at risk for use in homicide may be quite different from those most at risk for use in nonfatal assaultive violence. To begin to address this issue, it might be feasible to acquire information on firearms used in nonfatal shootings of law enforcement officers. From 1980 through 1989, 29 339 assaults on law enforcement officers with firearms were reported, of which 6496 (22%) involved injury.[10] The Department of Justice, the Consumer Product Safety Commission, and the Centers for Disease Control and Prevention have also explored adding firearm injuries to the existing National Electronic Injury Surveillance System.[19,20]

Concerns about generalizability do not pertain to the implications of these findings for law enforcement personnel themselves. This group is at high risk for occupational homicide, with rates of 20 to 25 homicides per 100 000 officers per year.[21,22] Most such deaths involve firearms; nearly half result from wounds to the front upper torso, and nearly 60% occur when the shooter and the victim officer are separated by 5 ft or less.[10] Our data show that 9-mm or .38-caliber handguns are most commonly involved. Law enforcement policymakers should consider requiring the use of soft "body armor," which provides protection against such weapons at close range.

Law enforcement officers also need enhanced safeguards against assaults using their own firearms. During 1980 to 1989, 21% of officers killed with a handgun were shot with their own service weapon, and others were killed with a handgun taken from a fellow officer. Law enforcement agencies often avoid purchasing handguns with trigger arresters or other safety devices in order to minimize the time delay to firing when the officer is in control of his or her own weapon.[23,24] These data suggest that attention should be given to modifications of firearm design that would not impede the officer's ability to use the weapon but would make it more difficult for others to do so.

The question of whether broad limitations on the availability of handguns would reduce the incidence of these events is outside the scope of this study. My data do indicate, however, that restrictions limited to imported firearms—a characteristic of much of American firearms policy—will not be effective. □

Acknowledgments

The author is grateful to Mr Thomas Wilson of the FBI's Uniform Crime Reporting Program and to Mr Jack Killorin, Chief, Public Affairs Branch, Bureau of Alcohol, Tobacco and Firearms, for furnishing the data used in this study. Dr Jay Beaumont, Prof Stephen Teret, and Ms Mona Wright provided helpful comments on earlier drafts. In addition, Ms Wright and Ms Barbara Claire provided expert technical assistance.

References

1. *Violent Crime in the United States.* Washington, DC: Bureau of Justice Statistics; 1991. Publication NCJ–127855.
2. *An Analysis of the Initial Year of Operation CUE in the Cities of Washington, DC, Boston, Mass, Chicago, Ill.* Washington, DC: Bureau of Alcohol, Tobacco and Firearms; 1977.
3. *Project Identification: A Study of Handguns Used in Crime.* Washington, DC: Bureau of Alcohol, Tobacco and Firearms; 1976. Publication ATF P 3310.1.
4. Brill S. *Firearm Abuse: A Research and Policy Report.* Washington, DC: Police Foundation; 1977.
5. Wright JD, Rossi PH, Daly K. *Under the Gun: Weapons, Crime, and Violence in America.* New York, NY: Aldine Publishing Co; 1983.
6. Burr DES. *Handgun Regulation.* Orlando, Fla: Florida Bureau of Criminal Justice Planning and Assistance; 1977. (Data presented in Wright JD et al. [ref 5].)
7. Wright JD, Rossi PH. *The Armed Criminal in America: A Survey of Incarcerated Felons.* Washington, DC: National Institute of Justice; 1985.
8. *Protecting America: The Effectiveness of the Federal Armed Career Criminal Statute.* Washington, DC: Bureau of Alcohol, Tobacco and Firearms; 1992.
9. Zimring FE. The medium is the message: firearm caliber as a determinant of death from assault. *J Legal Stud.* 1972;1:97–123.
10. *Law Enforcement Officers Killed and Assaulted.* Washington, DC: Federal Bureau of Investigation; 1989.
11. Mathews JH. *Firearms Identification.* Madison, Wis: University of Wisconsin; 1962:2.
12. Mathews JH. *Firearms Identification.* Springfield, Ill: Charles C Thomas; 1973:3.
13. Hill AB. *A Short Textbook of Medical Statistics.* 10th ed. Philadelphia, Pa: JB Lippincott Co; 1977.
14. Kahn HA, Sempos CT. *Statistical Methods in Epidemiology.* New York, NY: Oxford University Press; 1989.
15. Zimring FE. Street crime and new guns: some implications for firearms control. *J Crim Just.* 1976;4:95–107.
16. Katz D, Baptista J, Azen SP, Pike MC. Obtaining confidence intervals for the risk ratio in cohort studies. *Biometrics.* 1978;34:469–474.
17. Gardner MJ, Gardner SB, Winter PD. *Confidence Interval Analysis, Version 1.0* [computer program]. London, England: British Medical Journal; 1989.
18. Teret SP, Wintemute GJ, Beilenson P. The Firearm Fatality Reporting System—a proposal. *JAMA.* 1992;267:3073–3074.
19. *Description of Intentional Injury Special Study Conducted in 33 Selected NEISS Hospitals during the Treatment Period September 1–September 30, 1989.* Washington, DC: Consumer Product Safety Commission, Division of Hazard and Injury Data Systems; 1990.
20. Brick M, Tourangeau K, Cantor D. *A Statistical Evaluation and Cost Assessment of Using the National Electronic Injury Surveillance System (NEISS) to Obtain National Estimates of Nonfatal Firearm Injuries.* Rockville, Md: Westat Inc; 1992.
21. Kraus JF. Homicide while at work: persons, industries and occupations at high risk. *Am J Public Health.* 1987;77:1285–1289.
22. Davis H. Workplace homicides of Texas males. *Am J Public Health.* 1987;77:1290–1293.
23. Pagano CL. The 9mm auto as police firearm. *Police Marksman.* 1984;9:4–12.
24. Boyle JD, Uyeda M. Conversion to the Sig Sauer 9mm. *Law and Order.* 1988; 36(11):49–53.

The Prevalence and Correlates of Fear of Falling in Elderly Persons Living in the Community

ABSTRACT

Objectives. Fear of falling has been recognized as a potentially debilitating consequence of falling in elderly persons. However, the prevalence and the correlates of this fear are unknown.

Methods. Prevalence of fear of falling was calculated from the 1-year follow-up of an age- and gender-stratified random sample of community-dwelling elderly persons. Cross-sectional associations of fear of falling with quality of life, frailty, and falling were assessed.

Results. The prevalence of fear increased with age and was greater in women. After adjustment for age and gender, being moderately fearful of falling was associated with decreased satisfaction with life, increased frailty and depressed mood, and recent experience with falls. Being very fearful of falling was associated with all of the above plus decreased mobility and social activities.

Conclusions. Fear of falling is common in elderly persons and is associated with decreased quality of life, increased frailty, and recent experience with falls. (*Am J Public Health.* 1994;84:565–570)

Cynthia L. Arfken, PhD, Helen W. Lach, MSN, Stanley J. Birge, MD, and J. Philip Miller, AB

Introduction

Falling in elderly persons can lead to disability, hospitalizations, and premature death.[1,2] It may also result in a psychological trauma termed fear of falling.[3,4] Such fear may lead to staying home or other self-restriction of activities with debilitating physical consequences. Although fear of falling is recognized as potentially having both psychological and physical consequences, few studies have examined this fear.[5–11] Moreover, these studies have been limited by use of select samples: small groups of patients,[5,6] elderly persons living in housing units or retirement homes,[7–10] or volunteers.[11] Although these studies have reported isolated associations with factors such as decrease in mobility or poor balance, the association of fear of falling with decreased quality of life and increased frailty is unknown. Likewise, the prevalence of fear of falling in a random sample of community-dwelling elderly persons is unknown.

In addition to reporting associations with factors such as mobility or balance, two studies have examined the association of fear of falling with the prior falling experience of subjects.[7,8] One study examined prevalence of fear of falling in subjects with one or more falls in the prior year,[7] and the other examined restriction of activities after a noninjurious fall.[8] However, development of fear of falling may also be related to other measures previously used in research on geriatric falls: circumstances of the fall,[12,13] recurrent falls,[14–17] or severity of resulting injury.[18]

The current study was developed to examine prevalence of fear of falling and its association with measures of falling, quality of life, and frailty.

Methods

The current study was added at the 1-year follow-up of a cohort study of community-dwelling elderly persons. The primary purpose of the parent study was to design a simple screening tool to be used in the physician's office for predicting falls in elderly persons.

From the St. Louis OASIS (Older Adult Service and Information System) membership of 16 632 older adults, a gender- and age-stratified random sample of 1358 community-dwelling elderly persons was recruited (66% participation rate) to provide approximately equal numbers in four age groups: 65 through 69 years, 70 through 74 years, 75 through 79 years, and 80 years or older, with a ratio of 2 women to 1 man in each age group. Subjects from each stratum entered the study evenly over the course of a year with no excess of difficult-to-recruit subjects in the later months. To avoid subjects with severe cognitive impairment, individuals with 12 or more errors on the Mini-Mental State Examination[19] were excluded from the study (n = 15).

The fall surveillance procedure initiated after recruitment consisted of

Cynthia L. Arfken and J. Philip Miller are with the Division of Biostatistics, Washington University School of Medicine, St. Louis, Mo. Helen W. Lach and Stanley J. Birge are with the Program on Aging, Jewish Hospital, St. Louis, Mo.

Requests for reprints should be sent to Cynthia L. Arfken, PhD, Center of Health Behavior Research, 660 S Euclid, Box 8012, Washington University School of Medicine, St. Louis, MO 63110.

This paper was accepted July 13, 1993.

TABLE 1—The Prevalence (%) of Fear of Falling, by Gender and Age Groups

	Men				Women			
	66–70 y (n = 87)	71–75 y (n = 82)	76–80 y (n = 66)	81+y (n = 58)	66–70 y (n = 156)	71–75 y (n = 158)	76–80 y (n = 137)	81+y (n = 146)
No fear	86	85	86	79	79	63	60	55
Moderately fearful	10	15	12	16	15	26	28	34
Very fearful	3	0	2	5	6	11	12	11

subjects reporting falls to a hot line immediately after the fall plus monthly postcards to document any falls they had sustained, whether or not they had reported the falls by phone. Each month the investigators called all subjects who did not return postcards to verify that the subject was still in the study and solicit information about any unreported falls. After a report of a fall from the hot line, postcard, or phone, one of two trained gerontological nurses conducted a brief telephone interview with the subject to verify that a fall had occurred and the circumstances and outcome of it.[12]

One year after initial recruitment, 890 subjects received a structured in-home reassessment. Reasons for loss to follow-up included death (n = 30), moving (n = 2), refusal (n = 125), budgetary cuts (subjects recruited in the last 2 months were dropped [n = 290]), and incomplete information (incomplete interview [n = 33], telephone interview only [n = 69]). Those lost to follow-up were, on average, older than those for whom follow-up information was obtained. No other differences were found between those who were followed-up and those who were lost to follow-up. The final sample was predominately female (67%), White (93%), and educated (62% finished high school). Overall, 78% rated their health as excellent or good.

The in-home reassessment was conducted by lay interviewers and included demographic characteristics, health status, activity level, satisfaction with life, depressed mood, and a brief physical assessment. To measure the extent of fear of falling, the following question was asked of everyone: "At the present time, are you very fearful, somewhat fearful or not fearful that you may fall (again)?"

Quality-of-Life Measures

Quality-of-life measures covered mobility and social activities, satisfaction with life, and depressed mood. The mobility and social activities measures consisted of three questions on frequency of leaving the building or residence but not the yard, frequency of leaving the building and yard, and frequency of participating in social activities (i.e., "religious or club meeting, visiting friends or family, eating with other people, going to a social event or having friends in"). These three measures were dichotomized before this analysis by using a cutpoint of three times per week. These cutpoints were chosen for ease of interpretation and the criterion that at least 5% of the sample be in one of the categories. Satisfaction with life was dichotomized by using the contrast of "very" to "somewhat" or "not at all" satisfied. Depressed mood was analyzed with the Geriatric Depression Scale,[20] a 30-item scale where a higher score indicates a more depressed mood. Both the total score and a dichotomous score were used. The latter score used the recommended cutoff of 11 as suggestive of depression.[21]

Frailty Measures

Frailty measures covered episode of near fall, inability to walk 10 blocks, requiring assistance to climb stairs, poor vision limiting ambulation, use of assistive device to ambulate, general health status, and a balance score from the physical assessment. The balance assessment consisted of maintaining parallel, semitandem, and tandem stances for 10 seconds in each stance with eyes open and then with eyes closed. The balance score was computed as the number of stances correctly maintained (maximum = 6). This variable was dichotomized with a score of 4 or less suggesting impaired balance, a decision based on clinical interpretation.

Falls Measures

Measures of experience with falls covered a single fall, recurrent falls (two or more), a fall other than a trip or slip over a recognized hazard, a fall resulting in injury, a fall requiring medical treatment, a fall resulting in a fracture, or a fall with a delay in getting up (more than 1 minute). These falls must have occurred after recruitment and before the in-home reassessment.

Analysis

Statistical analysis consisted of first testing for differences in quality of life, frailty, and falling by the extent of fear of falling with the chi-square test for categorical variables and the Kruskal-Wallis Rank Test for the continuous score on the Geriatric Depression Scale. Then the associations of the categorical variables with fear of falling were summarized as odds ratios (ORs) obtained from multivariate logistic regressions with gender, age (as a continuous variable), and gender and age interaction terms. The corresponding large-sample 95% confidence intervals were also calculated.

Fear of falling in the logistic regressions was analyzed both as an ordinal variable with three levels and as a dichotomous variable. For the latter, we used models contrasting "moderately fearful" with "no fear" and models contrasting "very fearful" with "no fear." Logistic regression was also used to test the association of fear of falling as an ordinal variable with age and gender. Factors independently associated with fear of falling as an ordinal variable were identified by using logistic regression with stepwise selection after verifying that the proportional odds assumption was met. All analyses were conducted with SAS.[22]

To assess the effect of the chosen cutpoints, we used probability plots.[23] From these plots, the associations of mobility and social activities with fear of falling appeared monotonic. For balance, the associations would be maximized at cutpoints of 3 to 4.

Results

The majority of the sample (71%) expressed no fear of falling, whereas 9% reported they were very fearful of falling. Table 1 presents the fear-of-falling prevalences by gender and age groups. Women were more likely than men to express some fear of falling (35% vs 15%; $P < .0001$). However, fear of falling increased with age for both genders. In women, 21% of the youngest group (66–70 years) reported some fear of falling. This prevalence increased to 45% in the oldest age group (81+ years). In men, 14% of the youngest group reported some fear of falling, vs 21% of the oldest group.

Those who were very fearful of falling were most likely to have a decreased quality of life (Table 2). Almost half (48%) of those persons who were very fearful of falling stated that they were somewhat or not at all satisfied with life. In addition, 25% had scores suggesting depression. The continuous score on the Geriatric Depression Scale also differed by the extent of fear of falling: those who were very fearful had a mean score of 8.3 (SD = 4.8); those who were moderately fearful had a mean score of 7.2 (SD = 4.6); and those who were not fearful had a mean score of 4.8 (SD = 3.6) ($P < .0001$).

After gender and age adjustment, those who were very fearful of falling still had an elevated risk for worse quality of life on every measure (Table 3). This elevated risk was most pronounced for infrequently leaving the building (OR = 6.05). In contrast, those who were moderately fearful of falling did not have elevated risks for decreased mobility and social activities. However, they did have elevated risk for dissatisfaction with life (OR = 1.82) and depressed mood (OR = 2.20).

Those who were very fearful of falling were most likely to be frail or have falls on all measures (Table 4). Almost all (85%) of those who were very fearful of falling had impaired balance. The percentages of subjects with impaired balance did not materially change after excluding those who refused to perform the balance assessment (47% of subjects who had no fear, 66% who were moderately fearful, and 79% who were very fearful). Although the interviewers were instructed to stand near the subject in case assistance was needed, subjects fearful of falling may still have been fearful of trying the more difficult stances.

TABLE 2—The Prevalence (%) of Decreased Quality of Life, by Subject's Level of Fear of Falling

Quality-of-Life Measure	Subject's Level of Fear: No Fear (n = 633)	Moderately Fearful (n = 190)	Very Fearful (n = 67)	P[a]
Infrequently leave building but not yard[b]	3	3	13	<.0001
Infrequently leave building and yard[b]	6	10	22	<.0001
Infrequent social activities[b]	29	27	40	.11
Less than very satisfied with life[c]	21	34	48	<.0001
Depressed mood[d]	8	18	25	<.0001

[a]Chi-square test of homogeneity, $df = 2$.
[b]Less than three times per week.
[c]"Somewhat" or "not at all" vs "very" satisfied.
[d]Score of 11 or more on Geriatric Depression Scale.

TABLE 3—Associations of Fear of Falling with Quality-of-Life Measures

Quality-of-Life Measure	Moderately Fearful: Not Fearful OR	Moderately Fearful: Not Fearful 95% CI	Very Fearful: Not Fearful OR	Very Fearful: Not Fearful 95% CI
Infrequently leave building but not yard[a]	1.07	0.41, 2.80	6.05	2.46, 14.90
Infrequently leave building and yard[a]	1.00	0.54, 1.83	2.94	1.45, 5.96
Infrequent social activities[a]	0.90	0.62, 1.31	1.75	1.03, 2.97
Less than very satisfied with life[b]	1.82	1.26, 2.63	3.08	1.81, 5.25
Depressed mood[c]	2.20	1.36, 3.57	3.18	1.68, 6.04

Note. Odds ratios (ORs) and confidence intervals (CIs) were obtained from logistic regression models including gender, age, and gender and age interaction term.
[a]Less than three times per week.
[b]"Somewhat" or "not at all" vs "very" satisfied.
[c]Score of 11 or more on Geriatric Depression Scale.

Of those subjects who were very fearful of falling, 9% had had a fracture in the prior year compared with less than 0.5% of those with no fear and 2% of those with moderate fear. The reported fractures were ankle, hip, and leg (three subjects with no fear); pelvis, ribs, wrist, and shoulder (four subjects with moderate fear); and three hips, two shoulders, and one face (six subjects who were very fearful). Delay in getting up after a fall was similarly uncommon but more frequent in those very fearful of falling. Although this variable was dichotomized at greater than 1 minute, one subject lay on the ground over 1 hour after falling.

Almost all (91%) of those who were very fearful of falling reported at least one of the characteristics associated with frailty, in contrast to 59% of those with no fear of falling.

After gender and age adjustment, those who were very fearful of falling still had an elevated risk for both frailty and experience with falls on every measure (Table 5). This risk was most pronounced for fractures (OR = 28.16) and using an assistive device to ambulate (OR = 9.87).

To evaluate the independence of these associations, we conducted analysis using stepwise logistic regression models. Five variables were indepen dently associated with extent of fear of falling (Table 6). These factors were a fall resulting in a fracture (OR = 6.65),

TABLE 4—The Prevalence (%) of Frailty and Recent Experience with Falls, by Subject's Level of Fear of Falling

	Subject's Level of Fear			
	No Fear (n = 633)	Moderately Fearful (n = 190)	Very Fearful (n = 67)	P^a
Frailty				
Impaired balance[b]	51	70	85	<.0001
Episode of near fall	30	52	58	<.0001
Inability to walk 10 blocks	28	51	61	<.0001
Requiring assistance to climb stairs	28	62	67	<.0001
Vision limiting ambulation	3	5	16	<.0001
Fair or poor self-rated health	17	31	46	<.0001
Use of assistive device to ambulate	8	22	43	<.0001
Experience with falls[c]				
At least 1 fall	26	36	48	<.0001
Recurrent falls (2 or more)	8	13	22	<.0001
Fall besides trip or slip	11	17	34	<.0001
Fall with injury	11	14	24	.01
Fall with medical treatment	2	6	15	<.0001
Fall resulting in fracture	<0.5	2	9	<.0001
Delay in getting up after a fall, more than 1 min	3	4	15	<.0001

[a]Chi-square test of homogeneity, $df = 2$.
[b]Balance score of 4 or less.
[c]In the prior 12 months.

TABLE 5—Associations of Fear of Falling with Frailty and Recent Experience with Falls

	Moderately Fearful: Not Fearful		Very Fearful: Not Fearful	
	OR	95% CI	OR	95% CI
Frailty				
Impaired balance[a]	1.75	1.21, 2.55	4.44	2.15, 9.17
Episode of near fall	2.78	1.97, 3.93	3.61	2.12, 6.13
Inability to walk 10 blocks	2.21	1.55, 3.15	3.48	1.99, 6.08
Requiring assistance to climb stairs	3.65	2.54, 3.65	4.66	2.62, 8.30
Vision limiting ambulation	1.11	0.48, 2.60	4.79	2.09, 11.01
Fair or poor self-rated health	1.94	1.32, 2.85	3.72	2.16, 6.40
Use of assistive device to ambulate	2.82	1.74, 4.56	9.87	5.22, 18.68
Experience with falls[b]				
At least 1 fall	1.52	1.06, 2.17	2.49	1.48, 4.20
Recurrent falls (2 or more)	1.71	1.01, 2.89	3.12	1.61, 6.06
Fall besides trip or slip	1.47	0.92, 2.34	3.69	2.05, 6.63
Fall with injury	1.21	0.74, 1.98	2.26	1.20, 4.26
Fall with medical treatment	2.29	1.02, 5.18	6.31	2.59, 15.38
Fall resulting in fracture	4.23	0.86, 20.73	28.16	5.89, 134.66
Delay in getting up after a fall, more than 1 min	1.36	0.57, 3.23	4.52	1.94, 10.52

Note. Odds ratios (ORs) and confidence intervals (CIs) were obtained from logistic regression models including gender, age, and gender and age interaction term.
[a]Balance score of 4 or less.
[b]In the prior 12 months.

TABLE 6—Factors Independently Associated with Fear of Falling

	OR	95% CI
Fall resulting in fracture	6.65	2.02, 21.97
Requiring assistance to climb stairs	4.31	1.95, 9.54
Vision limiting ambulation	3.72	1.12, 12.34
Fair or poor self-rated health	2.89	1.32, 6.34
Fall besides trip or slip	2.71	1.29, 5.71

Note. Odds ratios (ORs) and confidence intervals (CIs) were obtained from stepwise logistic regression analysis.

requiring assistance to climb stairs (OR = 4.31), vision limiting ambulation (OR = 3.72), fair or poor self-rated health (OR = 2.89), and a fall besides a slip or trip (OR = 2.71).

Discussion

Fear of falling was common in this sample of relatively well community-dwelling elderly persons and comparable to that reported for a retirement center.[9] The associations of fear of falling with frailty and falls suggest a consistent trend toward a greater degree of association of these variables with a greater degree of fear. Although other unmeasured factors (such as personality and social factors) may play a role, the factors examined here had strong associations with fear of falling even after adjustment for gender and age. Moreover, in multivariate analysis, the extent of fear of falling was independently associated with experience with falls (fractures and falls not caused by a trip or slip) and frailty (problem climbing stairs, vision limiting ambulation, and poor health).

The strong and consistent association of fear of falling with decreased quality of life highlights the public health relevance. Given the design constraints, we cannot say whether the decreased quality of life preceded, accompanied, or followed the development of fear of falling. This caveat is also true for frailty and experience with falls. For this reason, we did not attempt more complex analyses assuming directionality (e.g., frailty leading to falling leading to

depressed mood leading to fear of falling). Asking about fear of falling at initial recruitment and at the reassessment would have been preferable both for determining generalizability and for predicting the development of fear of falling. However, the supposition that decreased quality of life follows a fall is supported by a case–control study[8] on noninjurious falls in retirement homes that found greater restriction of activities or death 180 days later.

Relying on the response to one question to determine the extent of fear of falling is a potential concern. Other researchers have defined fear of falling as curtailment of activities,[8] curtailment of activities because of fear of falling,[7] and self-efficacy in avoiding falls while performing specific activities.[11] These diverse definitions have advantages and disadvantages. Curtailment of activities has been criticized for lack of reliability.[11] Self-efficacy, while theoretically promising in being sensitive to change, suffers from difficulty in administration (i.e., subjects have to understand and estimate the extent of their confidence in their ability to avoid falls while performing a select list of activities). The approach we used is a step toward understanding this phenomenon and has been used by others.[9,10]

Three additional concerns are the low response rate for the parent study, the reduced cohort for the present study, and the reliance on self-report for frailty measures. The response rate, although similar to that reported by other investigators in gerontological research,[24] is troublesome because it is plausible that those who were fearful of falling might have been more likely than others to refuse to participate. This bias would have the effect of underestimating the prevalence of fear of falling. We interviewed subjects in their homes to minimize missing those subjects who were afraid to leave their home, but we may still have underestimated the prevalence of fear of falling.

Addressing the second concern, the cohort for the current study was 34% smaller than the recruited cohort, which limits generalizability. However, the majority of this decline (62%) was initiated by the investigators (i.e., terminating surveillance on those recruited in the last 2 months of the study). Analysis confirmed that the persons in the reduced cohort did not differ from those who were not followed, except that those who were not followed were older. This analysis suggests that we may have underestimated the prevalence of fear of falling in the original cohort and also limited our power to detect associations in the older age group.

The third concern is the reliance on self-report. The resulting associations between self-report of frailty and self-report of fear have competing interpretations. The subjects could have a rational fear of falling in response to their frailty. Alternatively, the subjects could be fearful and view themselves as more frail than they are. Although we included a performance-based measure of frailty (balance), the performance may be biased by fear of trying.[10]

One of the strengths of this study was that we used prospective documentation of falling with telephone verification for all the falls reported.[25] Therefore, we avoided biases due to potential differential recall of past falls. However, we did not assess falls occurring in the remote past or the impact of falls experienced by others, both of which could influence the development of fear of falling. For example, someone hearing of a traumatic fall leading to institutionalization may become fearful of falling. This analysis indicated that half of those who were very fearful of falling had not fallen in the prior 12 months. We don't know how long they had been fearful or how the fear was triggered. A more extensive look at the fear of falling should incorporate these factors.

In conclusion, we found fear of falling to be common in a sample of community-dwelling elderly persons. The extent of the fear was associated with self-reported and performance-based measures of frailty and prior experience with falling. The observed associations of fear of falling with decreased quality of life and mobility confirm the need for effective interventions to prevent and limit the consequences of falls in elderly persons. □

Acknowledgment

This work was supported by grant AG P01-6815 from the National Institute on Aging, National Institutes of Health.

References

1. Baker SP, Harvey AH. Fall injuries in the elderly. *Clin Geriatr Med.* 1985;1:501–512.
2. Committee on Trauma Research. *Injury in America.* Washington, DC: National Academy Press; 1985.
3. Tinetti ME, Speechley M. Prevention of falls among the elderly. *N Engl J Med.* 1989;320:1055–1059.
4. Hindmarsh JJ, Estes EH Jr. Falls in older persons. Causes and interventions. *Arch Intern Med.* 1989;149:2217–2222.
5. Murphy J, Isaacs B. The post-fall syndrome. A study of 36 elderly patients. *Gerontology.* 1982;28:265–270.
6. Bhala RP, O'Donnell J, Thoppil E. Ptophobia. Phobic fear of falling and its clinical management. *Phys Ther.* 1982;62: 187–190.
7. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *N Engl J Med.* 1988;319:1701–1707.
8. Vellas B, Cayla P, Bocquet H, de Pemille P, Albarede JL. Prospective study of restriction of activity in old people after falls. *Age Ageing.* 1987;16:189–193.
9. Walker JE, Howland J. Falls and fear of falling among elderly persons living in the community: occupational therapy interventions. *Am J Occup Ther.* 1991;45: 119–122.
10. Maki BE, Holliday PJ, Topper AK. Fear of falling and postural performance in the elderly. *J Gerontol.* 1991;46:M123–M131.
11. Tinetti ME, Richman D, Powell L. Falls efficacy as a measure of fear of falling. *J Gerontol.* 1990;45:P239–P243.
12. Lach HW, Reed AT, Arfken CL, et al. Falls in the elderly: reliability of a classification system. *J Am Geriatr Soc.* 1991;39:197–202.
13. Campbell AG, Borrie MJ, Spears GF. Risk factors for falls in a community-based prospective study of people 70 years and older. *J Gerontol.* 1989;44: M112–M117.
14. Campbell AJ, Reinken J, Allan BC, Martinez GS. Falls in old age: a study of frequency and related clinical factors. *Age Ageing.* 1981;10:264–270.
15. Nevitt MC, Cumming SR, Kidd S, Black D. Risk factors for recurrent nonsyncopal falls. A prospective study. *JAMA.* 1989; 261:2663–2668.
16. Cumming RG, Miller JP, Kelsey JL, et al. Medications and multiple falls in the elderly: the St. Louis OASIS Study. *Age Ageing.* 1991;20:455–461.
17. Dunn JE, Rudberg MA, Furner SE, Cassel CK. Mortality, disability, and falls in older persons: the role of underlying disease and disability. *Am J Public Health.* 1992;82:395–400.
18. Sattin RW, Lambert Huber DA, DeVito CA, et al. The incidence of fall injury events among the elderly in a defined population. *Am J Epidemiol.* 1990;131: 1028–1037.
19. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12: 189–198.
20. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res.* 1983;17:37–49.
21. Brink TA, Yesavage JA, Lum O, Heersema P, Adey M, Rose TL. Screening tests for geriatric depression. *Clin Gerontologist.* 1982;1:37–44.
22. *SAS Procedures Guide, Version 6, Third Edition.* Cary, NC: SAS Institute Inc; 1990.

23. Wartenberg D, Northridge M. Defining exposure in case-control studies: a new approach. *Am J Epidemiol.* 1991;133:1058–1071.

24. Kelsey JL, O'Brien LA, Griss JA, Hoffman S. Issues in carrying out epidemiologic research in the elderly. *Am J Epidemiol.* 1989;130:857–866.

25. Cumming SR, Nevitt MC, Kidd S. Forgetting falls. The limited accuracy of recall of falls in the elderly. *J Am Geriatr Soc.* 1988;36:613–616.

Upcoming Centers for Disease Control and Prevention Conferences

Call for Abstracts on Chronic Disease Prevention

The Centers for Disease Control and Prevention (CDC) is issuing a call for abstracts for the Ninth National Conference on Chronic Disease Prevention and Control, to be held in Washington, DC, December 7 through 9, 1994. Topic areas include chronic disease epidemiology, interventions, and intervention assessment and evaluation. The deadline for submission is *May 18, 1994.* For more information, contact Jack Friel, NCCDPHP, CDC, 4770 Buford Hwy NE, Mailstop K-30, Atlanta, GA 30341-3724; tel (404) 488-5390; fax (404) 488-5974 or (404) 488-5962.

BRFSS Conference to Be Held

CDC will sponsor the 11th annual Behavioral Risk Factor Surveillance System (BRFSS) Conference at the Atlanta Hilton and Towers in Atlanta, Georgia, June 8 through 10, 1994. The agenda will include utilization of behavioral risk factor data, issues and opportunities related to the analysis of data, updates on computer-assisted telephone interviewing systems, and recommendations to be adopted from an evaluation of the BRFSS. State BRFSS program coordinators, survey contractors, behavioral scientists, statisticians, epidemiologists, and others involved in the collection, analysis, and utilization of behavioral risk factor data are invited to attend. There is no registration fee. Additional information is available from the Behavioral Risk Factor Surveillance Branch, National Center for Chronic Disease, CDC, 4770 Buford Hwy NE, Mailstop K-30, Atlanta, GA 30341-3724; tel (404) 488-5292.

Using Physician Correspondence and Postcard Reminders to Promote Mammography Use

Stephen H. Taplin, MD, MPH, Carolyn Anderman, MPH, Lou Grothaus, MS, Susan Curry, PhD, and Daniel Montano, PhD

ABSTRACT

Objectives. In a health maintenance organization that mails letters to women recommending that they schedule mammograms, we conducted a randomized trial to evaluate simple methods of increasing the use of screening mammography.

Methods. Using a 2 × 2 factorial design, we tested the effects of (1) mailing the recommendation letter from each woman's primary care physician rather than from the program director and (2) sending a subsequent reminder postcard.

Results. Sending a reminder postcard nearly doubled the odds that women would get mammograms within 1 year (participate). The letter from the woman's personal physician had no effect. Attending a clinic more than 45 minutes from the screening center, being a current smoker, or being in fair or poor health were negatively associated with subsequently obtaining a mammogram. The odds of participation doubled if women had had previous mammograms.

Conclusions. When preceded by written recommendations to schedule mammograms, reminder postcards effectively increased participation. Future randomized trials to promote use of screening mammography should compare interventions with a reminder condition. (*Am J Public Health.* 1994;84:571–574)

Introduction

Breast cancer mortality remained constant from 1973 through 1988, partly because of the underuse of mammography.[1–7] By 1990, only about one third of women reported having had more than one mammogram.[6,7]

Mailing recommendations directly to women offers a promising approach to increasing mammography use. Such recommendations increased the use of blood pressure checks, influenza vaccines, and cancer screening tests.[8–19] In a Swedish randomized trial of mammography, mailed recommendations resulted in 92% of the women in the study getting mammograms.[20,21] In the United States, 21% to 65% of women obtain mammograms after mailed recommendations.[8,11,12,16,17,22]

We undertook a randomized trial to test whether we could increase participation in an established screening program that routinely mails recommendations to women due for mammograms.[12,23,24] We tested two enhancements: (1) mailing the recommendation letter from each woman's primary care physician rather than from the program director and (2) sending a subsequent reminder postcard. We based the interventions on the Health Belief Model, which suggests that cues to action promote preventive care behavior.[25,26]

Methods

Setting

Group Health Cooperative of Puget Sound is a staff model health maintenance organization with over 375 000 enrollees and the full complement of health care services. Adult Group Health Cooperative members choose from among family physicians or internists who each care for 1600 to 1800 enrollees. The populations of both the Group Health Cooperative and Puget Sound have a higher proportion of Caucasians (91%) than national figures (83%).[27,28] The Cooperative has a greater proportion of enrollees with more than 15 years of education (Group Health Cooperative = 34%, Puget Sound = 24%, and United States = 16%). There are also slightly fewer Cooperative enrollees with incomes below $15 000 (in 1984 dollars: Group Health Cooperative = 20%, Puget Sound = 23%, United States = 24%) and above $50 000 (in 1984 dollars: Group Health Cooperative = 13%, Puget Sound = 19%, and United States = 18%).[27,28]

Screening Program

In 1985, the Group Health Cooperative initiated a Breast Cancer Screening Program for women age 40 years and older.[12,23,24] Eighty-seven percent of women age ≥40 years complete an enrollment questionnaire that elicits breast cancer risk factors, perceived health status, smoking history, and previous mammography use. Once they are

Stephen H. Taplin is with the Department of Preventive Care and the Center for Health Studies, Group Health Cooperative of Puget Sound, Seattle, Wash. Carolyn Anderman, Lou Grothaus, and Susan Curry are with the Center for Health Studies, Group Health Cooperative of Puget Sound. At the time of the study, Daniel Montano was with the Department of Family Medicine, University of Washington, Seattle, Wash; he is now with the Battelle Institute, Seattle.

Requests for reprints should be sent to Stephen H. Taplin, MD, MPH, Preventive Care Research, Group Health Cooperative of Puget Sound, 1730 Minor Ave, Suite 1600, Seattle, WA 98101-1448.

This paper was accepted July 26, 1993.

TABLE 1—Baseline Characteristics, by Study Group

	Study Group				
	Control (n = 329)	Letter from Physician (n = 329)	Postcard Reminder (n = 335)	Both (n = 334)	*P*
Demographic					
Mean age, y	61.2	60.7	60.2	60.9	.39
Age ≥ 65 y, %	35	27	33	34	.17
Self-report of fair or poor health, %	14	13	14	15	.83
Behavioral influences					
Are current smokers, %	28	27	27	26	.95
History of breast cancer in aunt or grandmother,[a] %	4	3	4	4	.95
History of breast biopsy, %	1	3	2	2	.71
History of having mammograms, %	28	25	29	28	.64
Do breast self-examinations ≥ 12 times/y, %	28	24	24	26	.65
Logistic barriers					
Use clinic that is ≥ 45 min from BCSP[b] center, %	15	12	17	16	.26
Appointment wait time was ≥ 4 wk at time of invitation, %	53	43	48	48	.10

[a]Second-degree family history. Women with first-degree family history (mother, sister, daughter) of breast cancer were excluded from the study.
[b]BCSP = Breast Cancer Screening Program.

enrolled, an automated system sends letters to women due for mammograms. The letter is signed by the Breast Cancer Screening Program medical director, recommends scheduling a mammogram at one of three screening centers, and emphasizes the importance of the test.

Study Population and Design

We randomized women who were (1) age 50–79 years and had completed the questionnaire more than 1 year before randomization, (2) current Group Health Cooperative enrollees who had not been previously invited to a screening center, and (3) without a mammogram in the year before randomization. In order to have adequate numbers of women eligible for first invitations, we restricted the study to women who did not have a first-degree family history of breast cancer or more than one minor risk factor (early menarche, late menopause, second-degree family history of breast cancer).[24]

By making small changes in the Breast Cancer Screening Program correspondence, we tested the independent and combined effects of (1) having the recommendation letter come from the woman's physician rather than from the Breast Cancer Screening Program medical director (primary physician invitation) and (2) sending a reminder postcard to women 2 months after the recommendation letter (reminder postcard).

A sample of 1500 women was identified through the breast cancer program database, and these women were randomly allocated to four groups: (1) primary physician invitation (n = 329); (2) reminder postcard (n = 335); (3) primary physician invitation plus reminder postcard (n = 334); and (4) usual-care control group (n = 324). Subsequent to randomization, 11.5% of women were excluded from the study because they terminated Group Health Cooperative coverage (n = 34) or obtained a mammogram before being sent the recommendation letter (n = 135). Study data came exclusively from the program database and the risk factor questionnaire. All women were blinded to the study and received identical care to other women being invited at the same time.

Recommendation letters and reminder postcards were mailed directly to study women. When the primary care physician personally signed the letter, it also included a paragraph that emphasized the importance of the screening visit in the woman's overall health care. Reminder postcards were sent to all women in the appropriate study group without checking participation status beforehand.

Analysis

We compared baseline demographic characteristics, health status, screening history, and logistical barriers to participation among study groups. We dichotomized all characteristics (Table 1) and compared proportions among intervention populations using chi-square tests. We compared mean age and enrollment length using analysis of variance.

Using receipt of a mammogram within 12 months of the invitation letter (yes/no) as the dependent variable in a logistic regression model, we tested for the main effects of the primary physician letter, the reminder postcard, and the interaction of the physician letter and the reminder postcard. We then used chi-square tests with 1 *df* to compare each of the interventions with the control condition. These pairwise comparisons were repeated with a logistic regression model that included all the baseline characteristics. In addition, we fit five separate logistic models to test for these interactions: (1) physician letter with patient age and (2) reminder postcard with patient age, appointment wait time, travel time, or history of mammography.

Results

The groups did not differ significantly with respect to mean age (61 years) or average length of enrollment (10.4 years). As shown in Table 1, there were no significant differences among the four treatment groups on any of the baseline characteristics (Table 1).

The groups that received the reminder postcard had significantly greater rates of mammography use compared with control subjects (Table 2, $P < .001$). The odds of getting a mammogram increased at least 60% in the groups receiving a reminder postcard. The use of a letter from the woman's personal physician did not increase participation.

Adjusting for baseline covariates increased the odds ratio for both postcard groups to nearly 2 (Table 3). We did not find significant interactions between the physician letter and patient age, nor between the postcard reminder and the variables of age, appointment wait time, distance from the screening center, or having had a prior mammogram.

The covariate analysis also identified three baseline characteristics associated with a lower likelihood of obtaining a mammogram: reporting fair or poor health ($P = .02$), current cigarette smoking ($P = .0001$), and living more than 45 minutes from the screening center ($P = .0001$). Women who had had a previous mammogram were significantly more likely to participate ($P = .0001$). Because women had an average enrollment of 10 years, most of the previous mammograms occurred at the Group Health Cooperative.

Discussion

A letter from the woman's primary care physician did not increase the likelihood that she would get a mammogram within 1 year. In contrast, sending a subsequent reminder postcard nearly doubled the odds of participation.

McPhee and Detmer[29] recently demonstrated that mailed recommendations nearly doubled the odds that women would get a mammogram (62% vs 46%). Our study suggests that adding a reminder postcard 2 months after mailed recommendations would increase the rate even more.

Our results should encourage the evaluation of mailed reminders in economic or racial subgroups who were not represented in this study. Efforts to promote mammography use in Hispanic and Black populations currently emphasize encouraging physician recommendations.[30,31] However, a physician recommendation alone may not be sufficient to change behavior. Testing the effectiveness of reminders among these populations, once they receive the recommendation, would be a logical extension of our work.

The lack of effect of personal physician letters in this health maintenance organization setting may not generalize to the fee-for-service practice. In this study, the letter from the screening program medical director largely reflects the recommendation of the woman's physician. However, in fee-for-service practice, direct correspondence with women by a third party such as a radiology facility would be contrary to accepted notions of continuity. The role of continuity in achieving cancer screening goals needs closer evaluation, especially in fee-for-service practice.[32,33]

When considered in a national context, the results of this study suggest that future work to promote mammography must continue to be concentrated on women who have never obtained a mammogram. Our results demonstrate that the odds of getting a mammogram double if the woman has had a previous mammogram. Automated systems that mail reminders may be sufficient to encourage repeat mammography once a woman has had her first examination. Identifying and encouraging women who have not had the first mammogram should continue to be a priority.

Physician recommendations alone are not sufficient to ensure that women obtain a mammogram. Less than half the women who received that intervention participated, which is far less than the rates achieved in randomized trials.[20,22] More needs to be done to test methods of promoting participation. Interventions that address access barriers and patient characteristics should be compared with postcard reminders in future intervention trials.[34]

TABLE 2—Proportion of Study Group Women Who Obtained Mammograms within 1 Year

Intervention Group	Mammogram within 1 y of Recommendation, %	Odds Ratio[a]	95% Confidence Interval	P
Control[b] (n = 329)	46.8	1	...	
Primary physician invitation[c] (n = 329)	45.6	0.95	0.74, 1.36	.75
Postcard reminder[d] (n = 335)	58.5	1.60	1.18, 2.18	.003
Both[e] (n = 334)	61.7	1.83	1.34, 2.49	.0001

[a]Odds ratios were calculated from logistic regression model.
[b]All women received a letter signed by the program director that recommended they schedule a screening mammogram.
[c]The standard letters were modified to include the signature of the woman's physician.
[d]Women received reminder postcards subsequent to the recommendation letter.
[e]Women received both the letter from the woman's physician and the reminder postcard.

TABLE 3—Logistic Regression Analysis Comparing Odds of Obtaining Mammograms among Treatment Groups, With Baseline Characteristics Controlled

	Odds Ratio[a]	95% Confidence Interval	P
Interventions[b]			
Control	1.00	...	...
Primary physician invitation	0.98	0.69, 1.38	.89
Postcard reminder	1.92	1.36, 2.71	.0002
Physician letter + reminder	1.95	1.38, 2.74	.0001
Sociodemographics[c]			
If age ≥ 65 y	1.06	0.73, 1.53	.53
If have fair or poor health	0.63	0.45, 0.90	.02
Behavioral influences[c]			
If current smoker	0.48	0.37, 0.63	.0001
If second-degree family history of breast cancer	1.21	0.62, 2.33	.57
If history of breast biopsy	0.43	0.18, 1.05	.07
If history of mammograms	1.87	1.41, 2.48	.0001
If do breast self-examination ≥ 12 times/y	1.19	0.90, 1.58	.22
Logistic barriers			
If clinic was ≥ 45 min away	0.44	0.31, 0.62	.0001
If appointment wait was ≥ 4 wk	1.04	0.74, 1.47	.68

[a]From logistic regression calculated by using all variables and 1150 subjects.
[b]Coded so that control subjects are the reference group.
[c]Coded so that the reference group is the one without the characteristic.

A variety of surveys demonstrate that older women are least likely to get screening mammography.[7,35] Our work suggests that participation rates are similar across age categories if the recommendation is given. But the recommendation alone will not be sufficient to achieve high rates of participation among women of any age group.

Automated reminder systems applied to populations show promise as a technique for putting the potential mortality reductions afforded by screening mammography within reach. Improved interventions and the use of reminders in the fee-for-service setting need future evaluation. □

Acknowledgments

This study was made possible by grant CA34847 from the National Cancer Institute.

The authors acknowledge the careful attention to detail of Deb Timlin, Anne Howard, Nancy Snell, Jim Toomey, and Kathleen Hall. Their work and organization were greatly appreciated.

References

1. Shapiro S, Venet W, Strax P, Venet L. Current results of the breast cancer screening randomized trial: The Health Insurance Plan (HIP) of Greater New York Study. In: Day NE, Miller AB, eds. *Screening for Breast Cancer.* Toronto, Canada: Hans Huber Publishers; 1988.
2. Tabar L, Fagerberg G, Duffy SW, Day NE. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health.* 1989;43:107–114.
3. National Cancer Institute. Section III: mortality. In: Ries LAG, Hankey BV, Miller BA, et al., eds. *Cancer Statistics Review 1973–1988.* Bethesda, Md: National Institutes of Health; 1991;7. NIH Publichation 91-2789.
4. Howard J. Using mammography for cancer control: an unrealized potential. *Cancer.* 1987;37:33–48.
5. Centers for Disease Control. Provisional estimates from the National Health Interview Survey Supplement on Cancer Control—United States, January–March 1987. *MMWR.* 1988;37(27):417–419.
6. Lerman C, Rimer B, Trock B, Balshem A, Engstrom PF. Factors associated with repeat adherence to breast cancer screening. *Prev. Med.* 1990;19:279–290.
7. Centers for Disease Control. Use of mammography—United States, 1990. *MMWR.* 1990;39(36):621, 627–630.
8. McPhee SJ, Bird JA, Jenkins CNH, Fordham D. Promoting cancer screening. *Arch Intern Med.* 1989;149:1866–1872.
9. McDowell I, Newall C, Rosser W. A randomized trial of computerized reminders for blood pressure screening in primary care. *Med Care.* 1989;27:297–305.
10. Larson EB, Olsen E, Cole W, Shortell S. The relationship of health beliefs and a postcard reminder to influenza vaccination. *J Fam Pract.* 1979;8:1207–1211.
11. Myers RE, Engstrom PF, Rosan J, Amsel Z, Rimer B. Integrating breast cancer screening into an HMO medical care delivery system. *HMO Pract.* 1987;1(2): 67–74.
12. Thompson RS, Taplin SH, Carter AP, Schnitzer F. Cost effectiveness in program delivery. *Cancer.* 1989;64:2682–2689.
13. Larson EB, Bergman J, Heidrich F, Alvin BL, Schneeweiss R. Do postcard reminders improve influenza vaccination compliance? *Med Care.* 1982;20:639–648.
14. Brimberry R. Vaccination of high-risk patients with influenza. *J Fam Pract.* 1988;26:397–400.
15. Thompson RS, Michnich ME, Gray J, Friedlander L, Gilson B. Maximizing compliance with Hemoccult screening for colon cancer in clinical practice. *Med Care.* 1986;24:904–914.
16. Ornstein SM, Garr DR, Jenkins RG, Rust PF, Arnon A. Computer-generated physician and patient reminders: tools to improve population adherence to selected preventive services. *J Fam Pract.* 1991;32:82–90.
17. Clementz GL, Aldag JC, Gladfelter TT, Barclay AM, Brooks HF. A randomized study of cancer screening in a family practice setting using a recall model. *J Fam Pract.* 1990;30:537–541.
18. Banks NJ, Palmer RH. Clinical reminders in ambulatory care. *HMO Pract.* 1990;4(4):131–136.
19. Taplin SH, Anderman C, Grothaus L. Breast cancer risk and participation in mammographic screening. *Am J Public Health.* 1989;79:1494–1498.
20. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontott O. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am.* 1992;30:187–210.
21. Tabar L, Gad A, Holmberg L, Ljungquist U. Significant reduction in advanced breast cancer: results of the first seven years of mammography screening in Kopparberg, Sweden. *Diagn Imaging Clin.* 1985;54:158–164.
22. Fink R, Shapiro S. Significance of increased efforts to gain participation in screening for breast cancer. *Am J Prev Med.* 1990;6:34–41.
23. Carter AP, Thompson RS, Bourdeau RV, Andenes J, Mustin H, Straley H. A clinically effective breast cancer screening program can be cost-effective, too. *Prev Med.* 1987;16:19–34.
24. Taplin SH, Thompson RS, Schnitzer F, Anderman C, Immanuel V. Revisions in the risk-based breast cancer screening program at Group Health Cooperative. *Cancer.* 1990;66:812–818.
25. Becker MH, Haefner DP, Kasl SV, Kirscht JP, Malman LA, Rosenstock IM. Selected psychosocial models and correlates of individual health-related behaviors. *Med Care.* 1977;15(suppl):27–46.
26. Janz NK, Becker MH. The health belief model: a decade later. *Health Educ Q.* 1984;11(spring):1–47.
27. Pearson DC, Grothaus LC, Thompson RS, Wagner EH: Smokers and drinkers in a health maintenance organization population: lifestyles and health status. *Prev Med.* 1987;16:783–795.
28. *1980 Census of Population and Housing.* Washington, DC: US Dept of Commerce, Bureau of Census; 1980.
29. McPhee SJ, Detmer WM. Reminder interventions to improve delivery of cancer prevention services. Presented in part at the Annual Meeting of the American Society of Preventive Oncology. March 20, 1990; Bethesda, Md.
30. Fox SA, Stein JA. The effect of physician-patient communication on mammography utilization by different ethnic groups. *Med Care.* 1991;29:1065–1082.
31. Burack RC, Liang J. The acceptance and completion of mammography by older black women. *Am J Public Health.* 1989;79: 721–726.
32. Dietrich AJ, Marton KI. Does continuous care from a physician make a difference? *J Fam Pract.* 1982;15:929–937.
33. Kelly RB, Shank JC. Adherence to screening flexible sigmoidoscopy in asymptomatic patients. *Med Care.* 1992;30:1029–1042.
34. Wolosin RJ. Effect of appointment scheduling and reminder postcards on adherence to mammography recommendations. *J Fam Pract.* 1990;30:542–547.
35. The National Cancer Institute Breast Cancer Screening Consortium. Screening mammography: a missed clinical opportunity? *JAMA.* 1990;264:54–58.

The Characteristics of Northern Black Churches with Community Health Outreach Programs

Stephen B. Thomas, PhD, Sandra Crouse Quinn, PhD, Andrew Billingsley, PhD, and Cleopatra Caldwell, PhD

ABSTRACT

Objectives. The Black church has a long history of addressing unmet health and human service needs, yet few studies have examined characteristics of churches involved in health promotion.

Methods. Data obtained from a survey of 635 Black churches in the northern United States were examined. Univariate and multivariate statistical procedures identified eight characteristics associated with community health outreach programs: congregation size, denomination, church age, economic class of membership, ownership of church, number of paid clergy, presence of other paid staff, and education level of the minister.

Results. A logistic regression model identified church size and educational level of the minister as the strongest predictors of church-sponsored community health outreach. The model correctly classified 88% of churches that conduct outreach programs. Overall, the model correctly classified 76% of churches in the sample.

Conclusions. Results may be used by public health professionals and policy makers to enlist Black churches as an integral component for delivery of health promotion and disease prevention services needed to achieve the Year 2000 health objectives for all Americans. (*Am J Public Health.* 1994;84:575–579)

Introduction

Growing disparities in life expectancy and health status of Black Americans compared with Whites threaten the well-being, economic productivity, and social progress of our society as we approach the 21st century. Differences in the health status of Blacks and Whites have been documented in the United States as long as health data have been collected. These differences have persisted in spite of large increases in life expectancy and improvements in the health status of the general population. Health status is profoundly affected by poverty and racism. Poverty is the single most profound risk factor for disease.[1,2] Poverty with its concomitant lack of adequate housing, nutrition, and access to health care contributes greatly to the death toll on Black Americans. Failure to close the gap in health status between Whites and Blacks can also be directly attributed to lack of access to health care. George Lundberg, MD, editor of the *Journal of the American Medical Association,* attributes this lack of access to long-standing, institutionalized racial discrimination.[3]

One must be mindful that race of an individual is not an independent risk factor. In other words, Blacks do not have higher morbidity and mortality rates than Whites because they are Black. The landmark Black Report in Great Britain demonstrated that after decades of universal access to health services, health status inequalities between Whites and Blacks not only persisted, but were increasing.[1] In response to this evidence, William McBeath, Executive Director, American Public Health Association, stated that in America, "[W]e must assure access, not only of medical care, but also to a better standard of living—assuring nutritious food, basic education, safe water, decent housing, secure employment, and adequate income—these are the prerequisites of a healthy life-style."[1] In the absence of these prerequisites, the health status of Black Americans will continue to be cause for righteous indignation well into the 21st century.

In 1990, Dr Louis Sullivan, Secretary of Health and Human Services, presented to the nation *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.*[4] These objectives include special population targets for decreasing the disparity in health status and can provide a framework for health promotion and disease prevention activities conducted through the Black church. In this article we focus on results obtained from a survey of Black churches in the northern United States and describe how the Black church can contribute to the goal of achieving health for all Americans by the year 2000.

The Black church has historically been a major focus of the spiritual,

Stephen B. Thomas is with the Division of Behavioral Sciences and Health Education, Emory University School of Public Health, Atlanta, Ga. Sandra Crouse Quinn is with Westat, Inc, Rockville, Md. Andrew Billingsley is with the Department of Family and Community Development, University of Maryland, College Park, Md. Cleopatra Caldwell is with the Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, Mich.

Requests for reprints should be sent to Stephen B. Thomas, PhD, Division of Behavioral Sciences and Health Education, Emory University School of Public Health, 1599 Clifton Rd NE, Atlanta, GA 30329.

This paper was accepted July 7, 1993.

social, and political life of Black Americans. Billingsley and Caldwell[5] state that 84% of Black adults consider themselves to be religious and almost 70% are members of a church. Lincoln and Mamiya[6] estimate Black church membership at 24 million. It is also estimated that 65 000 to 75 000 Black churches of various denominations exist in the United States.[5]

The church has a long history of addressing unmet health and human service needs of the Black community.[7–9] As long ago as the 1920s, Mays and Nicholson[10] conducted a study of 609 urban churches and 185 rural churches. The authors found that community outreach programs included various activities aimed at addressing health needs, such as (1) programs to feed the unemployed, (2) free health clinics, (3) recreational activities, and (4) child care programs.

The existence of social networks and social support through the church provides the opportunity for health promotion programming.[8] Health programs that focus on behaviors that can be reinforced by social support are a natural arena for churches. Additionally, social and emotional support offered through church programs is perceived as an essential means to prevent the isolation that results from illness.

Historically, the Black church has been used by public health and medical professionals to gain access to those Blacks who are more difficult to reach through mainstream systems. The National Negro Health Movement, 1915 to 1950, represents the best historical example of how the Public Health Service utilized the church in a national strategy to bring modern public health practices to Blacks. The health status disparities that justified the need for a National Negro Health Movement are present today and demand that systematic research be conducted to determine how the vitality within Black churches may best contribute to reaching the objectives stated in *Healthy People 2000.*

Methods

The Black Church Family Project was a multiyear, national study of family-oriented, community outreach programs sponsored by Black churches. In 1979, the Sampling Section of the Survey Research Center of the Institute for Social Research at the University of Michigan designed a procedure to select the first nationally representative sample of Blacks for use in the National Survey of Black Americans. We assumed that the distribution of Black churches would reflect the distribution of Black households. The basic sampling frame included 76 primary sampling areas stratified by region. Data presented in this article are limited to a representative sample of 635 churches in the northern region of the United States that completed a computer-assisted telephone interview conducted during 1990 and 1991. Fifty-seven percent of the 1115 churches sampled completed the survey. Included in the category of nonrespondents were churches that were unreachable after three attempts and those churches that refused to participate.

Results

It must be noted that the Black church, like the Black community, is not homogeneous. Churches in the sample were drawn from a wide range of denominational affiliations. Baptist churches accounted for approximately 44% of the sample. Forty percent of these Baptist churches belonged to the National Baptist Convention—USA, Inc. Methodist churches represented 13% of the sample; of these, over 75% belonged to the African Methodist Episcopal, the African Methodist Episcopal Zion, and the Christian Methodist Episcopal denominations. Pentecostal churches represented 25.8% of the sample. Other churches—which included the Episcopalians, Presbyterians, the Church of God in Christ, Seventh Day Adventists, and smaller denominations—constituted 17.6%. Although the vast majority of churches (91%) had no paid female clergy, it is noteworthy that women constituted 50% to 100% of congregation members in 562 (88%) of the churches in the sample.

When senior ministers were questioned about the primary role of the contemporary Black church (membership service or community service), 5% emphasized serving the membership only and 9% indicated service to the community only, whereas the vast majority (86%) indicated that the role of the church was to serve church members and the community. The data revealed that 426 churches (67%) sponsored a total of 1804 community outreach programs. It was common to find multiple programs being operated by the same church. For example, 67% of churches operate one or more community outreach activities; 54% operate two or more community outreach activities; and almost 41% operate three or more community outreach activities.

An examination of the 1804 programs revealed that 51% were adult and family support programs, 31% were oriented to children and youth, 9% focused on community development, and 9% were programs for the elderly. Programs for children and youth addressed health issues including, but not limited to, human sexuality, acquired immunodeficiency syndrome (AIDS) education, drug abuse prevention, pregnancy prevention, and mental health.

Family programs focused heavily on the provision of basic human services like food, shelter, and child care. For those churches focusing on human service needs, the majority reported that participants were primarily low income. An examination of programs targeted to adults and families demonstrated a focus on counseling for AIDS risk reduction, drug abuse prevention, services to men and women in prison, and other health issues. Programs for the elderly focused on home health care, food delivery, medical care, and other basic human services. Furthermore, other health-related services included health clinics, screening for specific health problems, and health education workshops on various topics. Additionally, domestic violence programs for women and mental health programs for both youth and adults were sponsored. In detailed descriptions of 383 selected programs, we found that 94% of the programs providing basic human needs, counseling, and education were staffed primarily by volunteers.

An examination of data on the respondent's (typically the respondent was the senior minister) perceptions of community problems provides additional evidence that health issues are central to the mission of the church. Approximately 54% of ministers sampled considered drug abuse as the most serious problem in their community. Related programs included drug abuse counseling, AIDS support and prevention programs, drug education seminars, and sponsorship of Alcoholics and Narcotics Anonymous meetings.

The Black Church Family Project data also provide evidence that the churches sampled collaborate with secular service agencies and other churches. There were 450 churches (71%) identi-

TABLE 1—Ranking of Collaborations between Secular Agencies and 450 Black Churches

Secular Agency	% of Churches[a]
Local police	83
Local schools	82
Welfare departments	77
Hospitals	69
Local prisons	68
Health departments	65
Housing departments	65
Recreation departments	62
Youth organizations	56
Mental health departments	54
Employment agencies	51

[a]Percentages do not total 100 because respondents selected all categories that applied.

TABLE 2—Univariate Analysis of Church Characteristics Associated with Conduct of Community Health Outreach Programs

	No.	% with No Outreach	P
Membership size			<.01
Little (1–70)	146	60.3	
Small (71–175)	185	36.8	
Medium (176–400)	157	22.3	
Large (>401)	145	11.7	
Denomination			<.05
Baptist	277	38.3	
Pentecostal	164	35.4	
Methodist	82	23.2	
Church age			<.01
Young (1–40 y)	339	38.3	
Mid-age (41–75 y)	156	28.2	
Old (>75 y)	132	22.0	
Economic class			<.01
Working class	194	41.2	
Middle and working class	370	28.9	
Middle class	49	20.4	
Ownership of church			<.01
Rent	44	59.1	
Own	589	30.9	
Number of paid clergy			<.01
None	188	51.1	
One	352	27.6	
Two	59	13.6	
Three or more	31	19.4	
Other paid staff			<.01
No	305	43.6	
Yes	322	22.4	
Education level of minister			<.01
Less than high school	57	59.6	
High school	109	52.3	
Some college	93	33.3	
College degree	107	25.2	
Master's degree or above	188	16.5	

fied as having cooperative relationships with secular agencies. Table 1 identifies the types of agencies involved. Additionally, over 60% of churches collaborate with other churches in the same denomination as well as with churches of other denominations.

The existence of church programs that have collaborative relations with agencies such as welfare departments and that provide basic human services such as shelter and food indicates an ability to reach potentially underserved, poorly served, and never-served segments of the Black community. Collaborative relations with secular agencies may be strategically used to provide social support for health promotion and disease prevention initiatives designed to meet the Year 2000 health objectives.

Characteristics that influence participation in community outreach activities include church size, denomination, church age, economic class of membership, church ownership, number of paid clergy, existence of other paid staff, and education level of the minister (Table 2). Examination of data on church size demonstrates that a greater proportion of medium-size (78%) and large-size churches (88%) were involved in outreach programs.

Baptist, Pentecostal, and Methodist churches constituted 82% of the entire sample. Although a majority of these churches were actively involved in community outreach programs, some denominational differences were observed. Although there were more Baptist churches in the sample, a greater proportion of Methodist churches (77%) conducted outreach programs compared with Baptist churches (62%) and Pentecostal churches (65%).

The number of years the congregation had existed was an important factor. Approximately 78% of the churches over 75 years old reported involvement in community outreach, compared with 72% of churches between 41 and 75 years old and 62% of churches younger than 41 years old.

Economic class composition of the congregation was another important factor. Eighty percent of churches with a majority of members from the middle class sponsored community outreach programs. This figure declined to 71% among churches with congregations composed of members from the working and middle classes and dropped again to 59% among churches with congregations composed predominantly of working-class members.

Church ownership appeared to be another important factor. Overall, 70% of churches with paid mortgages engaged in community outreach programs. However, only 41% of church renters conducted outreach programs. This finding highlights the importance of the availability of church resources that may help to sustain and expand outreach programs. Additionally, ownership is related to a sense of independence and pride, both essential to the continuation of self-help and community health outreach.

The number of paid clergy was also associated with the conduct of community outreach programs. Only 49% of churches with no paid clergy were

TABLE 3—Logistic Regression Model of Church Characteristics Associated with Conduct of Community Health Outreach Programs

	Odds Ratio	95% Confidence Interval
Size		
Large (>401)	Referent	
Little (1–70)	0.11	0.04, 0.27
Small (71–175)	0.35	0.16, 0.76
Medium (176–400)	0.69	0.32, 1.50
Church denomination		
Methodist	Referent	
Baptist	0.64	0.30, 1.38
Pentecostal	1.80	0.76, 4.28
Church age		
Old (>75 y)	Referent	
Young (1–40 y)	1.32	0.63, 2.75
Mid-age (41–75 y)	1.07	0.51, 2.24
Economic class		
Middle class	Referent	
Working class	0.46	0.17, 1.21
Middle and working class	0.70	0.27, 1.79
Ownership of church		
Own	Referent	
Rent	0.67	0.28, 1.63
Number of paid clergy		
Two	Referent	
None	0.49	0.16, 1.49
One	0.83	0.31, 2.24
Three or more	1.26	0.24, 6.54
Other paid staff		
Yes	Referent	
No	0.99	0.56, 1.74
Education level of minister		
Master's degree or above	Referent	
Less than high school	0.32	0.13, 0.78
High school	0.29	0.14, 0.60
Some college	0.66	0.30, 1.42
College degree	0.90	0.44, 1.86

involved in outreach. For those churches with two paid clergy, over 86% offered outreach programs.

In addition to clergy, other paid staff are human resources who also contributed toward greater involvement in community outreach. Overall, 56% of churches with no additional paid staff engaged in outreach, whereas 78% of churches with paid staff conducted outreach programs.

Finally, it is noteworthy that the minister's level of education was strongly associated with church outreach activities. For example, only 40% of those churches whose minister had less than a high school education offered outreach programs compared with 83% of those churches whose ministers had a graduate degree.

The eight variables described in Table 2 were entered simultaneously into a logistic regression model. In this multivariate model (Table 3), church size and education level of the minister remained independent predictors of church outreach activities. Based on these variables, the logistic regression model correctly classified 88% of churches that conduct community health outreach programs and 51% of churches that did not participate in outreach activities. Overall, the model correctly classified 76% of churches in the sample.

Discussion

In summary, this study identified eight church characteristics that may be used by public health professionals to identify churches that conduct community health outreach programs. When these characteristics were analyzed in a multivariant model, church size and education level of the minister proved to be the strongest predictors of church-sponsored community health outreach programs.

Although the institution of the Black church has traditionally been sustained largely by volunteers, this study suggests that the availability of paid clergy and other paid staff to provide consistent leadership to various programs is a necessary factor that will determine the extent to which the church can continue the provision of community health outreach programs. Additionally, health professionals who belong to churches could provide a critical source of expertise and could be mobilized as volunteers to guide and implement community health programs. Although women comprise the majority of church members, they remain underrepresented in the ranks of paid clergy. However, their contribution as volunteers in sustaining church programs must be acknowledged.

Some may argue that the church may be unable to reach those populations most adversely affected by pressing public health problems such as tuberculosis, human immunodeficiency virus (HIV), violence, and substance abuse. However, results of our study demonstrate that Black churches work in collaboration with secular agencies such as prisons, local police departments, welfare departments, and public housing offices. This finding suggests that churches are already involved in delivery of services to poorly served, underserved, and never-served segments of our society. Unfortunately, there has been little systematic evaluation of the quality and efficacy of these efforts, nor have they been acknowledged in the public health literature.

Many churches struggle with moral issues related to the sexual and drug behaviors at the root of health problems such as substance abuse, HIV disease, and violence. For example, substantial opposition to needle exchange and condom distribution programs has been generated by Black churches in various cities.[11] In such cases involving controversial issues that conflict with church tenets, secular agencies may provide appropriate resources. This can free the church to focus on services that are more congruent with its mission.

This study identified characteristics associated with community health outreach that may be used by public health professionals to identify those Black churches most likely to be supportive of efforts to expand health promotion and disease prevention activities. Public health professionals should collaborate

with these churches by helping them focus community health outreach programs on specific *Healthy People 2000* objectives relevant to the target population and within the capability and scope of church resources and expertise.

It is noteworthy that graduate education of the minister is strongly associated with community health outreach. The Black clergy has historically provided social, political, and religious leadership to the Black community. Therefore, joint degree programs between schools of theology and schools of public health could train a cadre of ministers who could expand their sphere of influence to provide leadership on public health issues.

Finally, closing the health status gap between Black and White Americans will require public health professionals to draw on the strength, commitment, and credibility of the Black church supported by an adequate infrastructure. As the movement for national health care reform continues to grow, it will become necessary for public health professionals and policymakers to incorporate the Black church as an integral component for delivery of health promotion and disease prevention services needed to achieve the Year 2000 health objectives for all Americans.[11,12] □

Acknowledgments

The Black Church Family Project was supported by grants from the Ford Foundation and the Lily Endowment to Andrew Billingsley, Department of Family and Community Development, University of Maryland at College Park.

References

1. McBeath W. Health for all: a public health vision. *Am J Public Health.* 1991;81: 1560–1565.
2. National Research Council. *A Common Destiny: Blacks and American Society.* Washington, DC: National Academy of Sciences; 1989.
3. Lundberg G. National health care reform: an aura of inevitability is upon us. *JAMA.* 1991;265:2566–2567.
4. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50212.
5. Billingsley A, Caldwell C. The church, the family, and the school in the African American community. *J Negro Educ* 1991; 60:427–440.
6. Lincoln C, Mamiya L. *The Black Church in the African American Experience.* Durham, NC: Duke University Press; 1990.
7. Wiist W, Flack J. A church-based cholesterol education program. *Public Health Rep.* 1990;105:381–387.
8. Eng E, Hatch J, Callan A. Institutionalizing social support through the church and into the community. *Health Educ Q.* 1985;12:81–92.
9. Levin J. The role of the black church in community medicine. *JAMA.* 1984;76:477–483.
10. Mays B, Nicholson J. *The Negro's Church.* New York: Russell & Russell; 1933.
11. Thomas S, Quinn S. The burdens of race and history on black Americans' attitudes toward needle exchange policy to prevent HIV disease. *J Public Health Policy.* 1993;14:320–347.
12. Thomas S. Community health advocacy for racial and ethnic minority populations in the United States: issues and challenges for health education. *Health Educ Q.* 1990;17:15–19.
13. Thomas S. Health status of the black community in the 21st century: a futuristic perspective. In: Braithwaite R, Taylor S, eds. *Health Issues in the Black Community.* San Francisco, Calif: Jossey-Bass; 1992:338–349.

The Impact of the Safe Kids/Healthy Neighborhoods Injury Prevention Program in Harlem, 1988 through 1991

Leslie L. Davidson, MD, MSc, Maureen S. Durkin, PhD, DrPh, Louise Kuhn, MA, MPH, Patricia O'Connor, PhD, Barbara Barlow, MD, MS, and Margaret C. Heagarty, MD

ABSTRACT

Objectives. This study evaluated the effectiveness of a community coalition to prevent severe injuries to children in Central Harlem, New York, NY. It was hypothesized that injury incidence rates would decline during the intervention (1989 through 1991) relative to preintervention years (1983 through 1988); that the decline would be greatest for the targeted age group (5 through 16 years) and targeted injury causes (traffic accidents, assaults, firearms, outdoor falls); and that the decline would occur in the intervention community rather than a control community.

Methods. Surveillance of injuries that result in hospitalization and/or death among children in the two communities has been under way since 1983. Data from this surveillance were used to test whether the incidence of severe injury declined during the intervention; other temporal variations were controlled by Poisson regression.

Results. The incidence of injury among school-aged children in central Harlem declined during the intervention. The decline was specific to the targeted age group and targeted causes. A nonspecific decline also occurred in the control community.

Conclusions. The declining incidence rate in Central Harlem is consistent with a favorable program effect, but additional investigation of possible secular trend or spillover effects is needed. (*Am J Public Health.* 1994;84:580–586)

Introduction

Injury prevention has become a major public health goal in the United States[1] and community-based programs are widely recommended as a key strategy for the prevention of injuries,[2–4] but such programs are seldom evaluated rigorously. Because of variations in patterns of injury rates among communities, local surveillance systems must not only provide direction for the development of community-based programs but must also monitor and evaluate these programs.[4,5] In this report we present an evaluation of the effects on the incidence of severe injury of a community-based program, the Safe Kids/Healthy Neighborhoods Coalition, in Central Harlem, New York City.

Central Harlem health district is a disadvantaged community in New York City with a predominantly non-Hispanic Black population. It is contiguous to Washington Heights health district, a largely Hispanic community, which is slightly less disadvantaged. In 1990, 39.5% of Central Harlem residents lived below the poverty level, compared with 30.7% of Washington Heights residents and 19.3% of city residents as a whole (Table 1). Together the two communities, which constitute Northern Manhattan, contain 5.9% of New York City residents younger than 17 years.

The Northern Manhattan Injury Surveillance System (NMISS) was established at the two major hospitals serving these communities covering the years from 1983 onward. This system monitors the incidence of severe pediatric injuries (defined as those resulting in hospitalization or death). Preintervention surveillance data for the years 1983 through 1987 revealed three patterns: (1) the incidence of severe injuries was twice as high in Central Harlem as in Washington Heights; (2) the incidence was increasing among school-aged children (those aged 5 through 16 years) while declining among younger children (newborn through 4 years); and (3) the leading causes of severe injuries were falls and motor vehicle collisions. Assault was the cause of 9% of injury hospitalizations and 36% of injury deaths. Gunshot wounds were a major cause of fatal injuries (14%) to children in Harlem[6] and in New York City as a whole.[7]

Safe Kids/Healthy Neighborhoods Coalition

Parents and educators in Central Harlem requested a program in playground safety from health professionals. Surveys of the playgrounds revealed that they were sites for drug dealing and were in poor repair. Children in these play areas had very little adult supervision. The absence of safe play areas and

At the time of the study, Leslie L. Davidson was with the Gertrude H. Sergievsky Center and Division of Epidemiology, Columbia University, and the Department of Pediatrics, Harlem Hospital Center, New York, NY. Maureen S. Durkin and Louise Kuhn are with the Gertrude H. Sergievsky Center and Division of Epidemiology, Columbia University. Patricia O'Connor is with the Department of Psychology, Russell Sage College, Troy, NY. Barbara Barlow is with the Division of Pediatric Surgery and Margaret C. Heagarty is with the Department of Pediatrics, Harlem Hospital Center, New York, NY.

Requests for reprints should be sent to Leslie L. Davidson, MD, MSc, South East Institute of Public Health, David Salomons Estate, Broomhill Rd, Tunbridge Wells, Kent TN3 0TG, United Kingdom.

This paper was accepted November 4, 1993.

Editor's Note. See related editorial by Pless (p 537) in this issue.

supervised play activities for children was hypothesized to contribute to the high rates of outdoor injuries and assaults found among children in Harlem.

On the basis of these observations, at the end of 1988 the Harlem Hospital Injury Prevention Program, in collaboration with a variety of community groups and city agencies, initiated the Safe Kids/Healthy Neighborhoods Coalition, the initial goal of which was to reduce outdoor injuries. The Coalition soon grew to include assaults to school-aged children and developed alliances with city and community agencies and the private sector. Specifically, the Coalition worked to (1) renovate Central Harlem playgrounds; (2) involve children and adolescents in safe, supervised activities that would teach them useful skills (dance, art, sports, horticulture, carpentry); (3) provide injury and violence prevention education; and (4) provide safety equipment (bicycle helmets) at reasonable cost.

The development of the Safe Kids/Healthy Neighborhoods injury prevention program is described more fully elsewhere.[9] Over the first 3 years (1989 through 1991), 26 organizations (departments of city agencies, voluntary organizations, citizen groups) participated in the intervention. (Participation is defined as attending two or more planning meetings in any 1 year.) Injury prevention programs and activities were developed steadily over the period and involved large numbers of community children. The Department of Parks undertook repair of all playgrounds, including major capital improvements to five parks and playgrounds, and involved over 100 children in mural painting (three murals were completed). The Department of Transportation initiated an intensive pedestrian safety program that has reached all third-grade pupils in Harlem since 1990; the department has conducted other traffic safety education as well. The Harlem Hospital Injury Prevention Program initiated a dance program, an art studio, a Little League program, a winter baseball clinic, and a soccer league that have involved over 1000 children. Through bicycle safety programs, more than 500 bicycle helmets have been made available to the community.

TABLE 1—Demographic Characteristics of Central Harlem, Washington Heights, and New York City

	Central Harlem	Washington Heights	New York City
No. children aged <17 y			
1980	26 818	57 637	...
1990	28 457	66 305	...
Change, %	+5.8	+13.1	...
No. children aged <5 y			
1980	6 945	19 141	...
1990	9 203	21 770	...
Change, %	+24.5	+12.1	...
No. children aged 5–16 y			
1980	19 873	38 498	...
1990	19 254	44 535	...
Change, %	−3.2	+13.6	...
From 1990 census			
Non-Hispanic Black, %	87.6	18.8	25.2
Hispanic, %	10.2	63.4	24.4
Non-Hispanic White, %	1.4	15.1	43.2
Below poverty level, %	39.5	30.7	19.3
Receiving public assistance, %	29.6	23.7	13.1
Living two or more persons per room, %	1.6	3.5	1.8

Source. Data are from the 1980 and 1990 US censuses.[11,12]

Methods

Injury Hospitalizations and Deaths

Severe injury is defined as an injury resulting in hospitalization and/or death and refers to the *International Classification of Diseases* (ninth revision) external cause of injury codes (E-codes) between 800 and 999, which incorporate all trauma, poisonings, and burns that result from intentional, unintentional, and undetermined external causes.[10] Complications of hospitalization (E870–879) and therapeutic overdose (E930–949) are excluded from the definition. The study was restricted to residents of Northern Manhattan younger than age 17 (the criterion for admission to pediatric wards at one of the study hospitals) who were injured during the years 1983 through 1991.

Injury data for hospitalized cases were abstracted from the medical records of the two primary hospitals serving the area. Over the 9 years of the study, the data were abstracted by two physicians and a graduate student trained by an injury epidemiologist and pediatric surgeon. E-codes abstracted from the medical records were compared with E-codes on the discharge summary. Discrepancies were resolved in consultation with the injury epidemiologist. Pilot studies of data abstraction procedures showed high agreement between reviewers and data were routinely checked for accuracy.

Data on fatal injuries were obtained from New York City death certificate files. Reported deaths that occurred following admission to one of the two hospitals were verified in the data from the hospital records. The injury deaths among children who lived in the study area but who had not been admitted to one of the study hospitals were added. NMISS consists of the injury hospitalization and death data collected in this manner.

Adjustment for Incomplete Study Hospital Coverage

A comparison between NMISS data and the New York State Uniform Hospital Discharge Data (UHDD) revealed that 76% of all child residents of Northern Manhattan hospitalized for injury were admitted to one of the two study hospitals during the years 1983 through 1988. Since UHDD did not include the cause of injury until 1990, UHDD could not be used for the analysis directly. Instead, UHDD were used to adjust the incidence rates to compensate for the degree to which the data from the two hospitals underestimate the true incidence. This was done in the following manner. Within UHDD, the proportion of all injury-related hospitalizations in the study population admitted to one of the two study hospitals was calculated separately for each health district and for each year. Injury incidence in NMISS was adjusted by this proportion in each area and each year

TABLE 2—Numbers of Injuries and Mean Annual Adjusted Incidence Rates for All Causes of Severe Injury, by Age Group and Area

	Central Harlem		Washington Heights	
	Newborn–4 y	5–16 y	Newborn–4 y	5–16 y
No. injuries (1983–1991)				
Nonfatal	680	1217	975	1516
Fatal	29	37	34	14
Mean annual adjusted incidence rates[a]				
All severe injuries (1983–1991)	1251	968	611	468
Injury mortality (1983–1991)	37.9	21.2	17.6	11.4
All severe injuries, preintervention (1983–1988)	1382	1035	655	509
All severe injuries, intervention period (1989–1991)	989	833	523	387

[a]Adjusted rates (per 100 000 population) were calculated with numerators from the Northern Manhattan Injury Surveillance System, adjusted by the proportion estimated from Uniform Hospital Discharge Data to be hospitalized in other hospitals, and weighted averages of the population from the 1980 and 1990 US censuses.

separately. In 1989, 1990, and 1991 an average proportion for the preceding 6 years was substituted because UHDD for those years were found to be incomplete. The adjustment assumes that the cases treated in hospitals other than the two study hospitals were distributed by external cause in the same way as those in the database from the two hospitals. A comparison of the age, sex, racial, and broad diagnostic distributions of cases in the two data sets for 1983 through 1988, when UHDD were complete, show no major differences and suggest that this assumption is valid.[6] Distribution by E-code could not be compared because this information was not included in UHDD until the end of the study period.

Population Denominators

The US census (1980 and 1990) provided the population denominators for the two areas and two age groups (newborn through 4 years and 5 through 16 years).[11,12] A weighted average for each intercensus year was calculated for each area and age group. For example, the population estimate for the Central Harlem 0- through 4-year-old age group in 1984 was estimated as 60% of the 1980 count plus 40% of the 1990 count for the same area and age group.

Table 1 summarizes the demographic characteristics of the two communities from the 1980 and 1990 censuses. Substantial increases in the population occurred in Central Harlem in the younger age group and in Washington Heights in both age groups. Increases in birth rates in the two health districts during the past decade corroborate these census figures. Between 1980 and 1990, no striking changes occurred in the composition of the two health districts with respect to ethnic and economic indicators.

Statistical Analysis

The study design involves comparisons between the incidence rates of severe injuries during the intervention period (1989 through 1991) and those during the preintervention period (1983 through 1988). Three sets of comparisons are presented: (1) 5- through 16-year-old children vs newborn through 4-year-old children in Central Harlem (i.e., the rates in the targeted age group are compared with those in the nontargeted age group within the area in which the program was implemented); (2) 5- through 16-year-old children in Central Harlem vs children of the same age in Washington Heights (i.e., the rates in the area in which the program was implemented are compared with those in the neighboring community in which no specific injury prevention had occurred within the targeted age group); and (3) targeted vs nontargeted injuries among 5- through 16-year-old children in Central Harlem and in Washington Heights. Targeted injuries included all injuries related to vehicles, outdoor falls, assaults, and guns, regardless of intent. Nontargeted injuries included all others (poisonings, ingestions, burns, etc.). The three comparisons are used to strengthen inferences about the effect of the intervention by replicating the findings using different approaches to the data.

Incidence rates (adjusted as described above) for all injuries and for targeted and nontargeted injuries separately were calculated in 36 quarterly (3-month) intervals over the 9-year period for each age category (5 through 16 years and newborn through 4 years) and each area (Central Harlem and Washington Heights). Date of death or admission was used to indicate the quarter in which the injury occurred. Each case was assigned to a census tract, and therefore to either the Central Harlem or the Washington Heights health district, on the basis of the child's residential address.

The method of Poisson regression was used. The dependent variable was the injury rate in each quarter, which was modeled as a function of year (coded as a continuous variable), to account for net secular trends; season (coded as a four-way categorical variable), to account for marked increases in injury in the summer quarter; and intervention period (coded as a categorical variable corresponding to the time period before and after the development of the intervention). Poisson regression models can be understood as a subclass of generalized linear models in which the distribution of the error is assumed to be Poisson, the systematic effects are multiplicative, and the link function is log linear.[13–18] In each age and area category separately, Poisson regression models for rates of all injury and targeted and nontargeted injury types were fitted by means of maximum likelihood estimation and the statistical package GLIM.[19]

Poisson regression models have the useful feature of allowing the regression parameters to be interpreted as the log of the relative risk, adjusted for the other predictors in the model. In this case, the log of the regression parameter for the intervention period indicated the relative risk of injury following the intervention to that before the intervention, with annual and seasonal variations in injury incidence controlled. If, for example, incidence rates were decreasing prior to the intervention, then only if the decrease in the injury rate following the intervention was greater than this antecedent secular trend would the relative risk from the regression model indicate some protective effect of the intervention.

Results

Descriptive Epidemiology

Rates of severe injury hospitalization and mortality in the study population over the 9-year period of surveillance are displayed in Table 2 by age group and area. The injury rates for all severe injuries were approximately twice as high in Central Harlem as in Washington Heights for both age groups and declined for both areas and age groups over time.

Time Trends and Poisson Regression Analysis

Central Harlem: Comparison of incidence rates for the targeted and nontargeted age groups. The adjusted annual incidence rates of injuries in Central Harlem are plotted in Figure 1 for the two age groups. The rates can be seen to vary widely from year to year and appear to be decreasing over time in both age groups.

Poisson regression results show a significant decrease during the intervention period in the targeted age group (relative risk [RR] = 0.74, 95% confidence interval [CI] = 0.62, 0.89; Table 3). The intervention period is thus associated with a 26% reduction in the overall injury rate during the 3 years after the initiation of the prevention program compared with the 6 years before. In the younger, nontargeted age group, by contrast, no significant reduction in incidence occurred (RR = 1.06, 95% CI = 0.83, 1.35; Table 3). These relative risks are adjusted for annual and seasonal trends independent of the intervention. The model for children aged 5 through 16 years incorporates a marked seasonal effect but no significant annual trend. For younger children, seasonal trends are less consistent but the coefficient for annual trend is significant, reflecting a decrease in incidence with each sequential year.

School-aged children: Comparison of incidence rates for the targeted and nontargeted areas. The adjusted annual rates of injury from all causes in Central Harlem and Washington Heights for the age category 5 through 16 years are displayed in Figure 2. There appears to be a reduction in both areas in the last 3 years in this age category. The rates are substantially higher in Central Harlem than in Washington Heights in all years. Both areas show variability from year to year. The regression models show a

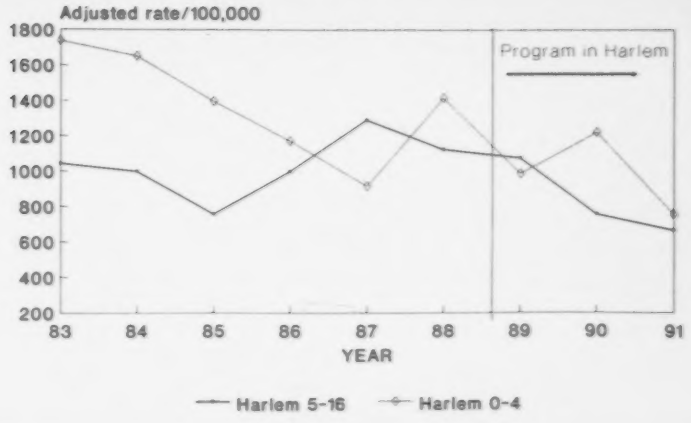


FIGURE 1—Adjusted annual rates per 100 000 population for all injuries in Central Harlem to children newborn through 4 years and 5 through 16 years.

TABLE 3—Results of Poisson Regression Models for Adjusted[a] Rates of All Injury Types: Effect of Intervention Period, with Year and Season Controlled

Model	β	SE (β)	P	Relative Risk	95% Confidence Interval
Central Harlem, children aged 5–16 y					
Year	0.02	0.016	.211		
Season					
2nd quarter	0.41	0.070	<.001		
3rd quarter	0.42	0.070	<.001		
4th quarter	−0.037	0.077	.631		
Intervention period	−0.30	0.090	<.001	0.74	0.62, 0.89
Central Harlem, children newborn–4 y					
Year	−0.08	0.022	<.001		
Season					
2nd quarter	0.10	0.095	.294		
3rd quarter	0.25	0.092	.007		
4th quarter	0.13	0.095	.171		
Intervention period	0.056	0.123	.460	1.06	0.83, 1.35
Washington Heights, children aged 5–16 y					
Year	0.019	0.016	.234		
Season					
2nd quarter	0.37	0.066	<.001		
3rd quarter	0.30	0.067	<.001		
4th quarter	−0.22	0.076	.004		
Intervention period	−0.36	0.087	<.001	0.70	0.59, 0.83
Washington Heights, children newborn–4 y					
Year	−0.039	0.020	.047		
Season					
2nd quarter	0.004	0.083	.960		
3rd quarter	−0.02	0.083	.844		
4th quarter	−0.06	0.084	.449		
Intervention period	−0.05	0.111	.680	0.96	0.77, 1.19

[a]Adjusted rates (per 100 000 population) were calculated with numerators from the Northern Manhattan Injury Surveillance System, adjusted by the proportion estimated from Uniform Hospital Discharge Data to be hospitalized in other hospitals, and weighted averages of the population from the 1980 and 1990 US censuses.

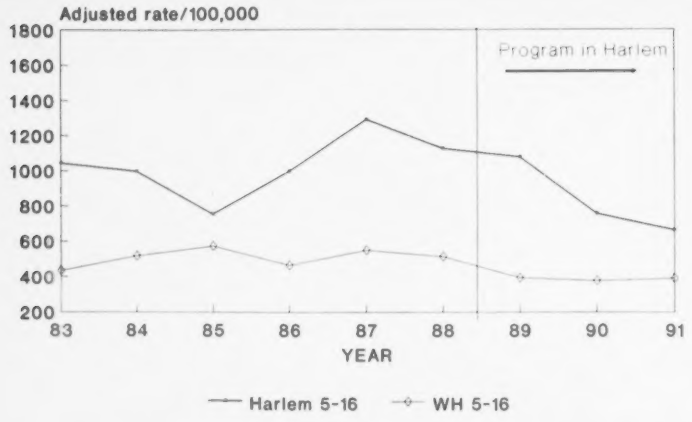
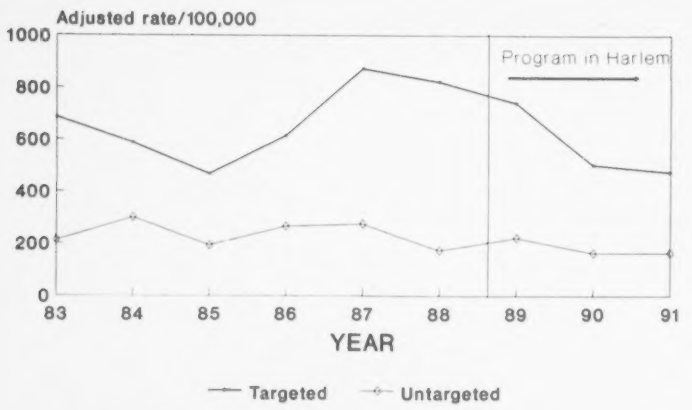


FIGURE 2—Adjusted annual rates per 100 000 population for all injuries in Central Harlem and Washington Heights (WH) to children aged 5 through 16 years.



Note. Targeted injuries are injuries related to motor vehicles, outdoor falls, guns, and assaults.

FIGURE 3—Adjusted annual rates per 100 000 population for targeted and nontargeted injuries in Central Harlem to children aged 5 through 16 years.

significant reduction for 5- through 16-year-olds in Washington Heights (RR = 0.70, 95% CI = 0.59, 0.83; Table 3). This reduction is slightly greater than the reduction observed in Central Harlem in the same age group. In the younger age group in Washington Heights, there is no significant reduction (RR = 0.96, 95% CI = 0.77, 1.19).

Central Harlem and Washington Heights: Comparison of incidence trends for targeted and nontargeted injuries. The adjusted annual incidence rates of targeted and nontargeted injuries among children aged 5 through 16 years in Central Harlem are displayed in Figure 3. The regression models show that in Central Harlem the decline during the intervention period is differential, affecting targeted injuries (RR = 0.56, 95% CI = 0.45, 0.71) but not other injuries (RR = 1.03, 95% CI = 0.78, 1.35; Table 4). Within the targeted injury category in Central Harlem, motor vehicle and assault injuries decreased significantly while outdoor falls did not. Figure 4 displays the adjusted annual incidence rates of targeted and nontargeted injuries among children aged 5 through 16 years in Washington Heights. In Washington Heights, the decline in incidence during the intervention period occurred only for motor vehicle injuries (Table 4).

Discussion

Few published evaluations of community-based injury prevention programs have used changes in injury rates as outcome measures. The Massachusetts statewide injury prevention program demonstrated a reduction in incidence in one of its targeted injury subtypes, motor vehicle occupant injuries.[20] In Sweden, a 4-year grassroots community-based injury prevention program showed postintervention decreases in home injuries (27%), occupational injuries (28%), and motor vehicle injuries (28%).[21] Both studies used a comparison between pre- and intraintervention injury rates in study and control communities. A study in the US Indian Health Service used regression analysis to demonstrate a program effect on changes in hospitalization rates for injury.[22]

This evaluation of the Harlem Safe Kids/Healthy Neighborhoods Coalition provides findings that are consistent with the hypothesized favorable effect of the program in reducing the incidence of severe injuries. It also produced findings that detract from the hypothesized effect. In favor of the hypothesis is the estimated 44% reduction in injury risk for targeted injuries in school-aged children in Central Harlem during the intervention period relative to the preintervention period. Moreover, the decrease within Central Harlem was specific to the age group (5 through 16 years) and largely specific to the injury types targeted by the prevention program. A finding not consistent with the study hypothesis is the significant decline in incidence also observed in the control area, Washington Heights (an estimated 30% reduction), for all severe injuries to school-aged children. The decline in Washington Heights occurred for both all targeted and all nontargeted causes, and, within the targeted category, it was restricted to motor vehicle injuries. Our conclusion from these mixed findings is that because the decline in incidence was specific to targeted injuries only in the intervention area, there is cause to be optimistic

about the effectiveness of the prevention program. At the same time, the significant reduction in motor vehicle injuries (one of the targeted causes) in both the intervention and control areas weakens the evidence in favor of the program's effectiveness in preventing this leading cause of injury.

The study also demonstrates the need for caution in using routinely collected hospitalization data for evaluation: had we used only the New York State UHDD for the follow-up period and not collected local surveillance data from hospital records, we would not have known that the UHDD were incomplete. This would have biased our results in favor of the hypothesis, for we would have found a spuriously low rate in the Harlem area compared with the Washington Heights area, for which the UHDD were more complete. The incomplete UHDD means that changes in referral patterns in the period since the intervention cannot be assessed. The adjustment makes the assumption that there were no such changes.

The observed decline in motor vehicle injuries, which for Northern Manhattan residents are predominantly pedestrian injuries, could be due to (1) a general decline in motor vehicle injuries, independent of any effects of the Safe Kids/Healthy Neighborhood Injury Prevention Program, or (2) a program effect that "spilled over" to the control area. In support of the first of these alternatives is a general national decline in motor vehicle death rates since 1989.[23] In support of the second alternative, preliminary data on the residence of program participants demonstrate that an increasingly large proportion (29%) of all participants in the program between 1989 and 1991 lived in Washington Heights. Further research is needed to investigate these hypotheses.

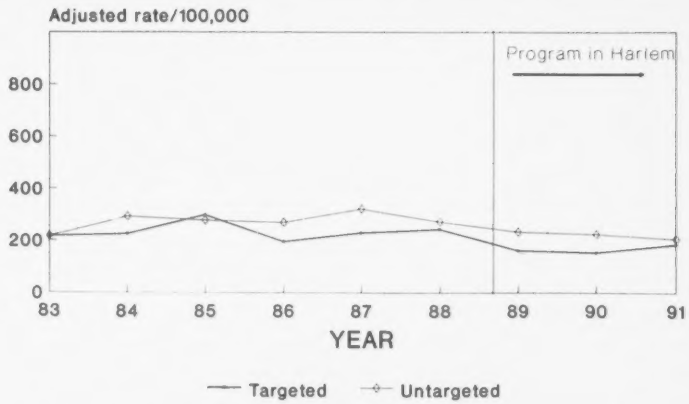
Another finding that is difficult to explain is the absence of a significant change in the incidence of outdoor fall injuries, a major cause of injury targeted by the program. One possible explanation for this finding is that the program activities are not effective in preventing outdoor fall injuries. The program contained no specific elements that addressed the problem of falls; rather, it focused on general improvements in environmental hazards (e.g., repair of dangerous fences and the provision of safer equipment) and increased supervision. The provision of many new supervised sports activities in the area may have actually resulted in an increased exposure to the risk of outdoor falls. Another possibility is that our ability to distinguish outdoor from indoor fall injuries was not adequate and that our measure of outdoor fall injuries was, therefore, not sufficiently valid to detect a change due to the intervention. The E-codes used to classify causes of injury do not always distinguish outdoor from indoor falls and data on location of injury event are incomplete. Cases for which the information was incomplete for classifying the location of falls were included in the untargeted group.

Despite these limitations, this study demonstrates the usefulness of injury surveillance both for guiding the development of a community-based and locally relevant injury prevention program and for evaluating the impact of such a program on injury incidence. In 1987, NMISS showed a dramatic increase in assault injuries and high rates of motor vehicle–pedestrian and fall injuries among school-age children in Central Harlem. It was these data that gave impetus and direction to the Safe Kids/Healthy Neighborhoods Injury Prevention Program in Central Harlem, initiated in 1988 and still active in 1993. The same surveillance system is providing the necessary data to evaluate the effects of the prevention program activities on the

TABLE 4—Effect of Intervention, with Year and Season Controlled, on Targeted and Nontargeted Adjusted Quarterly Injury Rates among Central Harlem and Washington Heights Children Aged 5 through 16 Years

	Central Harlem			Washington Heights		
	No.	RR	95% CI	No.	RR	95% CI
All targeted injuries[a]	664	0.56	0.45, 0.71	683	0.68	0.52, 0.87
Assault	199	0.52	0.34, 0.79	136	1.13	0.63, 2.03
Motor vehicle	302	0.45	0.32, 0.64	375	0.47	0.34, 0.66
Outdoor falls	121	1.35	0.75, 2.42	138	0.89	0.49, 1.54
Guns	100	0.56	0.31, 1.02	67	1.58	0.68, 3.66
All nontargeted injuries[b]	580	1.03	0.78, 1.35	833	0.71	0.57, 0.90

Note. RR = relative risk; CI = confidence interval.
[a]Targeted injuries include all assault injuries and all injuries related to motor vehicles, outdoor falls, and guns, regardless of intent.
[b]Nontargeted injuries include all unintentional nontrauma injuries and unintentional trauma injuries other than injuries related to motor vehicles, outdoor falls, and guns.



Note. Targeted injuries are injuries related to motor vehicles, outdoor falls, guns, and assaults.

FIGURE 4—Adjusted annual rates per 100 000 population for targeted and nontargeted injuries in Washington Heights to children aged 5 through 16 years.

incidence of severe injuries to children in Harlem. Such an evaluation is critically important for future planning of the program and, to the extent that the program is found to be successful, for facilitating sustained interest and support.

A major strength of this study is the use of Poisson regression to analyze injury rates over time, including periods before and during the intervention, adjusting for trends in the incidence rates that existed prior to the start of the prevention program. A simpler approach would have been to compare the mean annual incidence rates during the intervention period with those during the preintervention period. For example, for children younger than 5 years in Central Harlem the mean annual incidence rate was 28.4% lower during the intervention period than before (Table 2). The Poisson regression analysis shows a 6% increase (not statistically significant) in incidence for this age group during the intervention time period, after controlling for the annual decline in rates that occurred over the whole 9-year period. The regression approach provides a better way to detect changes likely to be due to the intervention rather than those associated with temporal variations in incidence that were present prior to the intervention. The Poisson regression approach could not be used to evaluate the impact of the program on injury fatalities because there were too few cases to produce stable quarterly mortality rates. It may be possible to use the approach in the future, using a larger control area and adding more years of data.

Another strength of this study is the use of multiple comparisons to assess the extent to which an observed reduction in injury incidence is specific to the age group, the area, and the types of causes targeted by the prevention program. The fact that these multiple comparisons gave mixed results indicates that more research is needed before a conclusion can be made about the effectiveness of the Safe Kids/Healthy Neighborhoods Coalition and injury prevention program. A simpler strategy might have provided results easier to interpret. For example, a strategy using no comparisons between age groups, areas, or causal categories would give results fully consistent with the hypothesized effectiveness of the program; a strategy using only the area comparison would support the conclusion of no measurable effect. The results of multiple comparisons have alerted us to the complexity of the temporal trends in pediatric injury incidence rates in Northern Manhattan and may have prevented us from drawing incorrect conclusions.

This study is limited to a relatively brief period of observation (3 years) following the start of the prevention program. Additional years of follow-up are needed to assess whether the decline persists and whether interventions only recently implemented have additional effect. Individual-level studies of associations between program involvement and exposure to program activities, on one hand, and injury incidence, on the other, will help to clarify the effect of the Safe Kids/Healthy Neighborhood Injury Prevention Program and its specific components on the risk of severe childhood injury in Central Harlem. These initial results are encouraging and suggest that the program is having a beneficial impact on the risk of severe injury to children in Harlem. □

Acknowledgments

This work was supported in part by grants from the Centers for Disease Control (R49/CCR 202585 and R48/CCR 205055), the Health of the Public (UCSF 1212SC), the Mellon Foundation, and the Robert Wood Johnson Foundation.

Thanks to Dr Jamal Naqvi for data abstraction over many years; to Drs Howard Andrews, Pat Shrout, Bruce Levin, and Elmer Struening for advice; to the New York City and State departments of health for making available the mortality and hospital discharge data; and to all the participants in the Harlem Hospital Injury Prevention Program and the Safe Kids/Healthy Neighborhoods Coalition for creating the intervention.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991:12. DHHS publication PHS 91-50212.
2. Baker S. Childhood injuries: the community approach to prevention. *J Public Health Policy.* 1981;2:235–246.
3. Micik SH, Miclette M. Injury prevention and the community: a systems approach. *Pediatr Clin North Am.* 1985;32:251–256.
4. The National Committee for Injury Prevention and Control. *Injury Prevention: Meeting the Challenge.* Oxford, England: Oxford University Press; 1989:23–24. (Published as a supplement [no. 3] to *Am J Prev Med* 1989, vol 5.)
5. Graitcer PL. The development of state and local injury surveillance systems. *J Safety Res.* 1987;18:191–198.
6. Davidson LL, Durkin MS, O'Connor P, Barlow B, Heagarty MC. The epidemiology of severe injuries to children in Northern Manhattan: methods and incidence rates. *Paediatr Perinat Epidemiol.* 1992;6:153–165.
7. Davidson LL. Violence to children and youths in urban communities. In: Schwarz DF, ed. *Report of the 23rd Ross Roundtable on Critical Approaches to Common Pediatric Problems.* Columbus, Ohio: Ross Laboratories; 1992:21–29.
8. Laraque D, Davidson LL, Barlow B, et al. Playground injuries: designing an intervention. In: Harris BH, Coran AG, eds. *Progress in Pediatric Trauma.* 3rd ed. Boston, Mass: Nob Hill; 1989:238. Abstract.
9. Davidson LL, Barlow B, Thompkins T, et al. Beyond the hospital's ivory tower: community injury prevention in Harlem. *Child.* 1994; in press.
10. *International Classification of Diseases, 9th revision: Clinical Modification.* Hyattsville, Md: National Center for Health Statistics; 1980. DHHS publication PHS 80-1260.
11. *1980 Census.* Washington, DC: US Dept of Commerce, Bureau of the Census; 1981. On-line census tables.
12. *1990 Census.* Washington, DC: US Dept of Commerce, Bureau of the Census; 1992. On-line census tables.
13. Clayton D, Schifflers E. Models for temporal variations in cancer rates, I: age-period and age-cohort models. *Stat Med.* 1987;6:449–467.
14. Clayton D, Schifflers E. Models for temporal variations in cancer rates, II: age-period-cohort models. *Stat Med.* 1987; 6:469–481.
15. Kleinbaum DG, Kupper LL, Muller KE. *Applied Regression Analysis and Other Multivariate Techniques.* Boston, Mass: PWS Kent; 1988.
16. McCullagh P, Nelder JA. *Generalized Linear Models.* New York, NY: Chapman-Hall; 1989.
17. Frome EL. The analysis of rates using Poisson regression models. *Biometrics.* 1983;39:665–674.
18. Kuhn L, Davidson LL, Durkin MS. A comparison of Poisson regression and time series analysis to evaluate the effects of an injury prevention program on child injury incidence. Submitted for publication.
19. Payne CD. *The GLIM System Release 3.77.* Oxford, England: Numerical Algorithms Group; 1986.
20. Guyer B, Gallagher SS, Chang B-H, Azzara CV, Cupples LA, Colton T. Prevention of childhood injuries: evaluation of the statewide childhood injury prevention program (SCIPP). *Am J Public Health.* 1989;79:1521–1527.
21. Schelp L. The role of organizations in community participation—prevention of accidental injuries in a rural Swedish municipality. *Soc Sci Med.* 1988;11:1087–1093.
22. Robertson LS. Community injury control programs of the Indian Health Service: an early assessment. *Public Health Rep.* 1986;101:632–637.
23. Technical Report, *1992 Traffic Crashes, Injuries, and Fatalities—Preliminary Report.* Washington, DC: US Dept of Transportation, National Highway and Transportation Safety Administration; March 1993.

Low-Income Neighborhoods and the Risk of Severe Pediatric Injury: A Small-Area Analysis in Northern Manhattan

Maureen S. Durkin, PhD, DrPH, Leslie L. Davidson, MD, MSc, Louise Kuhn, MA, MPH, Patricia O'Connor, PhD, and Barbara Barlow, MD, MS

ABSTRACT

Objectives. The purpose of this study was to investigate the relationship between socioeconomic disadvantage and the incidence of severe childhood injury.

Methods. Small-area analysis was used to examine socioeconomic risk factors for pediatric injury resulting in hospitalization or death in Northern Manhattan, New York, NY, during a 9-year period (1983 through 1991).

Results. The average annual incidence of all causes of severe pediatric injury was 72.5 per 10 000 children; the case-fatality rate was 2.6%. Census tract proportions of low-income households, single-parent families, non–high school graduates, and unemployment were significant predictors of risk for both unintentional and intentional injury. Among the socioeconomic factors considered, low income was the single most important predictor of all injuries; other socioeconomic variables were not independent contributors once income was included in the model. Compared with children living in areas with few low-income households, children in areas with predominantly low-income households were more than twice as likely to receive injuries from all causes and four and one half times as likely to receive assault injuries. The effect of neighborhood income disparities on injury risk persisted after race was controlled.

Conclusions. These results illuminate the impact of socioeconomic disparities on child health and point to the need for injury prevention efforts targeting low-income neighborhoods. (*Am J Public Health.* 1994;84:587–592)

Introduction

Childhood injuries are now recognized as a leading public health problem in the United States and other developed countries.[1–4] Studies of the incidence and risk factors of severe injuries, including nonfatal injuries, are needed to provide direction in planning and implementation of appropriate prevention programs.[5,6] Low socioeconomic status is one risk factor that is relatively well documented for injury fatality, particularly for house fires and homicide,[1,7–11] but it has rarely been investigated for the far more frequent outcome of severe nonfatal injury.[12] In general, the epidemiology of nonfatal injuries differs from and cannot be extrapolated from that of fatal injuries.[13] Another limitation of current knowledge is the absence of information on specific components of low socioeconomic status that affect a child's risk of injury.

We used methods of small-area analysis[14–17] to investigate associations at an ecological (census tract) level between multiple indicators of socioeconomic status and the incidence of severe pediatric injury in an inner-city environment in New York City. One aim of the study was to describe the variability between census tracts in socioeconomic indicators and in mean annual incidence rates during the 9-year study period. Another was to investigate the extent to which geographic variations in injury incidence are explained by variations in socioeconomic indicators. The results are presented for overall severe pediatric injury incidence rates and for broad subcategories of causes.

Methods

The study population consisted of residents of two Northern Manhattan health center districts, Central Harlem and Washington Heights, with a combined population in 1990 of 94 762 children younger than 17 years (Table 1).[18] According to major socioeconomic indicators, Northern Manhattan is disadvantaged relative to the rest of New York City, and the Central Harlem health district is disadvantaged relative to Washington Heights. For example, according to the 1990 census, 34% of the households in Northern Manhattan (40% in Central Harlem and 31% in Washington Heights) lived below the poverty level, compared with 19% in New York City as a whole.[18]

Data Sources

The Northern Manhattan Injury Surveillance System collected data on all injuries to residents of Northern Manhattan younger than age 17 that resulted in hospitalization or death during the years 1983 through 1991.[19] Data were abstracted from medical charts of the two

At the time of the study, Maureen S. Durkin, Leslie L. Davidson, and Louise Kuhn were with the Gertrude H. Sergievsky Center and the Division of Epidemiology, Columbia University, New York, NY. Patricia O'Connor is with the Department of Psychology, Russell Sage College, Troy, NY. Barbara Barlow is with the Department of Pediatric Surgery, Harlem Hospital Center, New York, NY.

Requests for reprints should be sent to Maureen S. Durkin, PhD, DrPH, Sergievsky Center, Faculty of Medicine, Columbia University, 630 W 168th St, New York, NY 10032.

This paper was accepted November 12, 1993.

Editor's Note. See related editorial by Pless (p 537) in this issue.

TABLE 1—Socioeconomic and Demographic Characteristics of Northern Manhattan

Total population	381 995
Children younger than 17 y	94 762
Race/ethnicity, %	
Hispanic	47.0
Non-Hispanic Black	40.2
Non-Hispanic White	10.7
Other	2.1
Age distribution, %	
0–4 y	8.1
5–9 y	7.3
10–14 y	6.8
15–16 y	2.6
>16 y	75.2
Socioeconomic indicators, %[a]	
Low income	33.1
Single parents	36.9
Non–high school graduates	45.0
Crowding	2.7
Mothers not employed	57.4
Unemployed men	15.6

Note. Northern Manhattan comprises two health center districts, Central Harlem and Washington Heights.
Source. Data are from the 1990 US census.[18]
[a]See text for definitions.

TABLE 2—Mean Annual Injury Incidence and Mortality (per 10 000 Population) in Northern Manhattan, 1983 through 1991

	Incidence	Mortality	Case–Fatality
All injuries	72.5	1.8	2.6
Unintentional injuries	57.8	0.8	1.4
Assault injuries	7.1	0.6	10.1
Self-inflicted injuries	2.0	0.1	3.2
Injuries of undetermined cause	4.7	0.3	7.1
All motor vehicle injuries	12.6	0.3	2.4
Pedestrian injuries	7.8	0.2	2.1
Occupant injuries	1.1	0.01	1.1
Fall injuries	16.4	0.2	1.0
Gunshot injuries	3.1	0.3	11.1
Burn injuries	9.5	0.3	3.1
By age group (all injuries)			
0–4 y	81.5	2.5	3.1
5–9 y	59.0	0.9	1.6
10–14 y	64.2	1.2	1.9
15–16 y	88.0	3.2	3.9
By sex			
Male	92.8	2.4	2.6
Female	50.5	1.2	2.4
By race/ethnicity			
Hispanic	44.3	0.7	1.7
Non-Hispanic Black	96.0	2.8	3.0
Non-Hispanic White	27.9	1.1	4.1

Note. Northern Manhattan comprises two health center districts, Central Harlem and Washington Heights.

study hospitals (Harlem Hospital Center and Columbia-Presbyterian Medical Center, the two major hospitals serving the area) and New York City death certificates. Data on residents of other areas hospitalized in these two hospitals were excluded. Injury cases include all those with *International Classification of Diseases* (9th revision) E-codes 800 to 999, which incorporate all trauma, poisonings, and burns resulting from unintentional, intentional, and undetermined external causes. Detailed data are included in the surveillance system for all deaths (n = 162, including 33 deaths following admission to one of the two study hospitals) and for the 4592 survivors who were hospitalized in the two study hospitals during the 9-year period. Less than 2% of the hospitalizations involved readmission of the same child for injury. According to a computerized hospital discharge database maintained by the New York State Department of Health, 76% of all Northern Manhattan children with nonfatal injuries were hospitalized in one of the two study hospitals (72% of those in Central Harlem and 85% of those in Washington Heights). The incidence rates given in Table 2 of this paper were first calculated on the basis of the detailed data on patients treated in the study hospitals, then inflated by 24% to represent the proportion of residents treated in other hospitals.[19] To validate this approach, we compared patients from the study area treated in other hospitals with those treated in the study hospitals and found no major differences in age, sex, ethnicity, or frequency of diagnoses. It was not possible to compare external cause (E-code) distributions because this information was not included in computerized New York State hospital discharge data prior to 1990. The deaths (assumed to be complete) were then added. The denominators (population younger than 17 years) of the incidence rates were weighted averages obtained from US census data for 1980 and 1990.[18,21]

Socioeconomic Variables

The following socioeconomic indicators for each census tract in the study area were derived from the 1990 census: (1) low income (proportion of households with annual incomes of less than $10 000); (2) poverty (proportion of households below the poverty level); (3) single parents (proportion of family households with only one parent present); (4) non–high school graduates (proportion of the population older than 17 years who did not graduate from high school); (5) crowding (proportion of occupied housing units with two or more persons per room); (6) mothers not employed (proportion of women with children younger than 6 years old who are not employed); and (7) unemployed men (proportion of men who are unemployed). The poverty variable is derived from the variables low income and household size; we found its relation to injury risk to be similar to, if somewhat weaker than, that of low income. For brevity's sake, the results for poverty are not reported here. The frequency in the study population of each socioeconomic variable examined, as well as age and ethnic variables, is given in Table 1. After reviewing socioeconomic indicators for Northern Manhattan from the 1980 and 1990 censuses and finding no major changes over the decade, we used only the indicators from the 1990 census in this paper.

Statistical Analysis

For the small-area analyses, the unit of analysis is the census tract. Injury data from the two study hospitals for residents of 75 of Northern Manhattan's 84 tracts were used. Three tracts with fewer than 100 children and six on the boundaries of the area with high proportions of injury admissions to other hospitals were excluded from the small-area analyses to avoid instability of rates due to small denominators. Coefficients of variation[22–24] were computed for each variable by dividing the standard deviation by the mean and chi-square tests of homogeneity were done to assess whether variation between census tracts in both independent and dependent variables was larger than would be expected by chance.[24]

Consistency of the injury rates within each tract over the 9-year period was assessed with Cronbach's alpha coefficient.[25,26] Regression analyses were done only for the overall and cause-specific injury rates considered consistent or reliable (alpha coefficients greater than .40). A simple linear regression of each social indicator with a given type of injury rate and a multiple linear regression of combinations of social indicators with a given injury rate were performed.[27] Injury incidence (all causes and major subtypes) was the dependent variable and indicators of socioeconomic status were the independent variables. In the multiple regression analyses, independent variables with P values of less than .10 were entered. The R^2 is the percentage of variance in injury incidence explained by a given socioeconomic variable after any other variables in the model are controlled.

To further assess low-income neighborhood as a risk factor, we computed the ratio of the incidence rate in census tracts with moderate and large numbers of low-income households to that in a reference area composed of tracts with few low-income households. Confidence intervals around the rate ratios were computed by the Taylor series approximation.[28]

Results

Average annual incidence rates for all injuries and those due to the most frequent external causes, by age, sex, and ethnicity, are displayed in Table 2. Measures of variation between census tracts in injury rates as well as socioeconomic indicators show large and significant variations across the area (Table 3). The bivariate regression results for all injuries show large and significant positive associations with the variables low income and single parents; smaller but still significant positive associations with non–high school graduates, mothers not employed, and unemployed men; and no significant association with crowding (Table 4). The same directions of associations are seen for each category of injury cause, although the strengths and significance levels of the associations vary for different categories (Table 4).

TABLE 3—Range and Variability of Census Tract Mean Injury Incidence and Socioeconomic Indicators, and Year-to-Year Consistency[a] of the Rank Order of Tract-Level Injury Incidence Rates

	Range of Means across Census Tracts	SD	Coefficient of Variation	Cronbach's α
Socioeconomic indicators[b]				
Low income	8.1–77.3	12.47	0.35	
Single parents	4.8–59.1	12.42	0.34	
Non–high school graduates	2.4–75.6	14.35	0.25	
Crowding	0.0–8.4	1.75	0.71	
Mothers not employed	0.0–100.0	17.14	0.32	
Unemployed men	0.0–55.1	9.67	0.57	
Mean injury incidence[c]				
All injuries	3.9–160.2	36.6	0.60	.88
Unintentional injuries	3.9–127.0	28.1	0.57	.83
Assault injuries	0.0–32.8	6.6	1.03	.73
Self-inflicted injuries	0.0–6.6	1.9	1.27	.17
Injuries of undetermined cause	0.0–13.3	3.3	0.84	.30
Motor vehicle injuries	0.0–36.8	7.6	0.73	.46
Pedestrian injuries	0.0–32.9	6.0	0.86	.58
Occupant injuries	0.0–4.6	1.0	1.36	.14
Fall injuries	0.0–39.1	9.3	0.62	.66
Gunshot injuries	0.0–13.4	3.3	1.23	.46
Burn injuries	0.0–276.4	62.5	0.77	.52

[a]As indicated by Cronbach's alpha.
[b]See text for definitions.
[c]Mean annual injury incidence per 10 000 population in each of the 75 included census tracts, 1983 through 1991.

The apparent protective effect of crowding, which is significant for assaults and burns and nearly significant for gunshot injuries, is unexpected and is probably due to confounding by ethnicity. The percentage of Hispanic residents in a census tract is positively correlated with crowding (Pearson correlation coefficient = .76), and Hispanic children have lower injury rates than non-Hispanic Black children. When racial and ethnic composition of neighborhoods is controlled, crowding is no longer associated with assault, burn, or gunshot injury risk.

In the multiple regression analyses, low income (or poverty) was the strongest predictor of all injuries and of each causal category examined. Changes in the order in which the variables were entered did not affect this result. Once income is in the model, crowding continues to have a small negative effect for assault, burn, and gunshot injuries, and the effects of the remaining socioeconomic variables are insignificant.

The rate ratio estimates show a strong positive association between residence in a low-income neighborhood and risk of severe pediatric injury (overall and for each causal category) (Table 5). Compared with children in neighborhoods with few low-income households, children living in largely low-income neighborhoods have more than twice the risk of severe injury from all causes and four and one half times the risk of severe assault injury. For most injury categories considered, a dose–response association is present (Table 5).

Race, ethnicity, and socioeconomic status are highly confounded in Northern Manhattan census tracts, making it difficult to demonstrate an effect of poverty independent of racial composi-

TABLE 4—Simple Linear Regression Results: Socioeconomic Indicators[a] and Injury Incidence[b]

	β[c]	P	R^2
All injuries			
Low income	16.713	.000	.325
Single parents	14.879	.000	.255
Non–high school graduates	9.727	.001	.146
Crowding	−30.972	.205	.022
Mothers not employed	7.522	.002	.124
Unemployed men	13.457	.002	.127
Unintentional injuries			
Low income	12.942	.000	.330
Single parents	11.354	.000	.252
Non–high school graduates	7.673	.000	.153
Crowding	−20.290	.280	.016
Mothers not employed	5.849	.002	.127
Unemployed men	10.312	.002	.126
Assault injuries			
Low income	2.821	.000	.280
Single parents	2.281	.000	.182
Non–high school graduates	1.374	.010	.088
Crowding	−10.043	.022	.070
Mothers not employed	1.206	.007	.097
Unemployed men	2.383	.002	.121
Motor vehicle injuries			
Low income	3.304	.000	.297
Single parents	3.151	.000	.268
Non–high school graduates	1.746	.004	.110
Crowding	−5.028	.320	.013
Mothers not employed	1.069	.036	.059
Unemployed men	2.343	.009	.090
Pedestrian injuries			
Low income	2.858	.000	.355
Single parents	2.240	.000	.217
Non–high school graduates	1.356	.004	.106
Crowding	−5.715	.151	.028
Mothers not employed	0.702	.083	.040
Unemployed men	1.710	.016	.077
Fall injuries			
Low income	3.469	.000	.215
Single parents	2.603	.002	.120
Non–high school graduates	2.246	.002	.119
Crowding	−5.914	.343	.012
Mothers not employed	1.879	.002	.119
Unemployed men	3.380	.002	.123
Gunshot injuries			
Low income	1.049	.001	.153
Single parents	1.070	.000	.158
Non–high school graduates	0.419	.070	.044
Crowding	−4.280	.053	.050
Mothers not employed	0.510	.023	.068
Unemployed men	0.531	.188	.024
Burn injuries			
Low income	2.246	.000	.201
Single parents	2.009	.000	.159
Non–high school graduates	1.452	.004	.111
Crowding	1.583	.706	.002
Mothers not employed	1.151	.006	.100
Unemployed men	0.847	.262	.017

[a]See text for definitions.
[b]Mean annual incidence per 10 000 population in each census tract, 1983 through 1991.
[c]The beta coefficient for low income can be interpreted as the change in injury incidence per 100 000 population for every 1% change in the percentage of low-income households in a census tract. The beta for single parents is the change in injury incidence per 100 000 population for every 1% change in the percentage of family households with only one parent present. The beta coefficients for the remaining independent variables can be interpreted similarly.

tion. In only 3 of 31 predominantly African-American tracts were fewer than one third of the households considered low-income. In an attempt to control for race, we computed incidence rates for non-Hispanic Black children only and found low-income neighborhood to be a significant predictor of injury risk within this group ($R^2 = .14$, $P < .001$). The marginally lower effect in this race-specific regression analysis compared with that of the total population is likely to be due to smaller variation in the proportions of low-income families in African-American neighborhoods. In a relative risk analysis of incidence of all injury types for Black children, the relative risk for low-income neighborhood is 1.9 (95% confidence interval = 1.5, 2.7), not significantly different from the effect among all children. We also tried controlling for race in a multiple regression analysis; after we entered census tract proportion of non-White residents ($R^2 = .49$, $P < .0001$, for all injuries), low income did not explain significant additional variation in injury risk (incremental R^2 for low income = .02, $P = .09$, for all injuries). Perhaps neighborhood racial composition is more accurately measured by the census and provides a better and more comprehensive marker of socioeconomic disadvantage than household income and other intended socioeconomic indicators. Thus, the confounding of poverty with race may be only partly extricable in the census measures; once race is held constant, our ability to show an effect of income is limited.

Discussion

These results show that children living in low-income neighborhoods are at increased risk of severe injury from both unintentional and intentional causes. This general finding is consistent with previous studies of fatal childhood injuries[7,9–11] but has not to our knowledge been demonstrated for the much more frequent occurrence of injury hospitalization. One study, a case–control study carried out in Belfast and published in 1959,[12] showed significant individual-level associations between several indicators of poverty and pediatric morbidity due to motor vehicle collisions.

Another case–control study of nonfatal child bicycle and pedestrian injuries found a twofold excess risk for low vs high maternal education but no signifi-

cant association with residence in an economically disadvantaged neighborhood.[29] Two other studies of nonfatal child pedestrian injuries found that more cases occurred in low-income areas than in more advantaged and less densely populated areas, but these studies did not compute population-based incidence rates to control for differences between areas in population size.[30,31] The present study, based on incidence rates and confined to an inner-city area, shows significant ecological associations between socioeconomic disadvantage and incidence of all major categories of severe pediatric injury.

The two analytic approaches, linear regression and rate ratio estimation, produced similar results except that the regression results imply that low-income neighborhood may be a slightly more important risk factor for unintentional injuries ($R^2 = .33$) than for assault injuries ($R^2 = .28$), while the rate ratio results suggest that low-income neighborhood is a much more important risk factor for assault injuries (rate ratio = 4.5) than for unintentional injuries (rate ratio = 2.0). Perhaps this is because the association between low-income neighborhood and injury risk is not linear for all causes, as assumed by the regression approach. For unintentional causes the association fits a linear model, whereas for assault injuries there appears to be a threshold effect. Limitations in the use of linear regression to estimate measures of effect, due to model mispecification and violations of assumptions regarding the distributions of the variables, have been noted.[32]

Because small-area analysis reveals ecological rather than individual-level associations, it is advantageous for identifying risk factors that operate on an ecological level.[33–35] For example, if poor neighborhoods have fewer safe play areas, more children are likely to play in the streets, abandoned buildings, and other hazardous areas. Poor neighborhoods may present other hazards, such as broken playground equipment, broken glass, poor housing, drug activity, violence, and a high prevalence of firearms in the hands of children, adolescents, and young adults.[36] In addition, children in low-income neighborhoods may have limited opportunities to engage in organized extracurricular activities. Other factors may operate at an individual or family level, such as supervision, knowledge, and education regarding child safety; access to safe and affordable child care; teenaged pregnancy; use of violence in response to conflicts; mental health; and stress. For risk factors that work at an individual level, the possibility of ecological fallacy[35,37] must be considered when the small-area approach is used.

TABLE 5—Injury Rate Ratios (RRs) in Moderately Low-Income and Largely Low-Income Census Tracts Relative to Rates in Referent Tracts

	Moderately Low-Income Tracts		Largely Low-Income Tracts	
	RR	95% CI	RR	95% CI
All injuries	1.7	1.6, 1.8	2.2	2.0, 2.4
Unintentional injuries	1.6	1.5, 1.8	2.0	1.8, 2.3
Assault injuries	2.6	1.9, 3.7	4.5	3.2, 6.5
Self-inflicted injuries	2.3	1.3, 4.1	1.9	1.0, 3.7
Injuries of undetermined cause	1.6	1.2, 2.2	1.8	1.2, 2.6
Motor vehicle injuries	1.9	1.6, 2.4	2.5	2.0, 3.2
Pedestrian injuries	2.0	1.5, 2.7	3.1	2.3, 4.2
Fall injuries	1.5	1.3, 1.8	1.9	1.5, 2.2
Gunshot injuries	2.4	1.5, 3.9	3.4	2.0, 5.7
Burn injuries	1.4	1.1, 1.8	1.6	1.3, 2.1

Note. Referent census tracts were those in the lower 25th percentile in proportion of low-income households (n = 19). Moderately low-income tracts = census tracts in the middle two quartiles in proportion of low-income households (n = 37); largely low-income tracts = census tracts above the 75th percentile in proportion of low-income households (n = 19). CI = confidence interval.

The proportions of variance in incidence rates explained by low income (0.15 to 0.36) appear modest relative to R^2s seen in small-area studies of other health outcomes.[14–17,22–24] The causes of injury are multifactorial and most cases may not be associated with economic disparities. Preventable injuries certainly occur to children across all social strata, although they appear to occur at a higher rate in poorer neighborhoods. An additional consideration is that a program to prevent childhood injuries was initiated in Central Harlem during the last 3 years covered by this study.[38] Methodological factors likely to have attenuated the observed effects are the presence of random measurement error in the variables of interest and the fact that coverage by the two study hospitals was lower in the more disadvantaged section of Northern Manhattan, Central Harlem (see Methods section), which also had the highest injury incidence. Although the rates were adjusted overall to take into account treatment in other hospitals, information was not available to make differential adjustments for incomplete case ascertainment by census tract.

Further research is needed to determine the extent to which the association between low-income neighborhood and pediatric injury risk is causal and to identify specific mechanisms and pathways. Prevention efforts, however, need not wait for further research. An immediate implication of the results presented here, along with findings on injury morbidity and mortality in other populations, is that programs, policies, and legislation for child injury prevention should target socioeconomically disadvantaged communities. Within low-income urban neighborhoods, efforts should focus on prevention of a wide range of intentional and unintentional causes of injury, including assaultive and suicidal behaviors, access to firearms, motor vehicle collisions (especially with pedestrians), falls, and burns. A model injury prevention program targeting many of these causes has been initiated in Central Harlem.[38] □

Acknowledgments

This research was supported by grants from the Centers for Disease Control (R49/CCR 202585 and R48/CCR 205055), the National Institute of Neurological Diseases and Stroke (R29 NS27971-03 and R29 NS27971-04), and the Robert Wood Johnson Foundation.

The authors would like to acknowledge Dr Margaret Heagarty for significant contributions to the design of the study and to data collection; Dr Jamal Naqvi for assistance with data collection for the Northern Manhattan Injury Surveillance System; Ms Louise Berenson of the New York City Department of Health for assistance with the death certificate data; and Drs Howard Andrews and Elmer Struening and an anonymous reviewer

for suggestions regarding the small-area analyses and content of the paper.

References

1. Baker SP, O'Neill B, Ginzburg MJ, Li GH. *The Injury Fact Book.* 2nd ed. New York, NY: Oxford University Press; 1992.
2. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50212.
3. Fingerhut L, Kleinman J. Trends and current status in childhood mortality, United States, 1900–1985. *Vital Health Stat [3].* 1989; no. 26:1–44.
4. *World Health Statistics Annual.* Geneva, Switzerland: World Health Organization; 1990.
5. Haddon W. Advances in the epidemiology of injuries as a basis for public policy. *Public Health Rep.* 1980;95:411–421.
6. Robertson L. *Injury Epidemiology.* New York, NY: Oxford University Press; 1992.
7. Wise PH, Kotelchuck M, Wilson ML, Mills M. Racial and socioeconomic disparities in childhood mortality. *N Engl J Med.* 1985;313:360–366.
8. Baker SP, Whitfield RA, O'Neill B. Geographic variations in mortality from motor vehicle crashes. *N Engl J Med.* 1987;316:1384–1387.
9. Mierley MC, Baker SP. Fatal house fires in an urban population. *JAMA.* 1983;249: 1466–1468.
10. Nersesian WS, Petit MR, Sharper R, Lemieux D, Naor E. Childhood death and poverty: a study of all childhood deaths in Maine, 1976 to 1980. *Pediatrics.* 1985;75:41–50.
11. Winpisinger KM, Hopkins RS, Indian RW, Hostetler JR. Risk factors for childhood homicides in Ohio: a birth certificate–based case–control study. *Am J Public Health.* 1991;81:1052–1054.
12. Backett EM, Johnston AM. Social patterns of road accidents to children: some characteristics of vulnerable families. *BMJ.* 1959;1:409–412.
13. The National Committee for Injury Prevention and Control. Injury prevention: meeting the challenge. *Am J Prev Med.* 1989;5(suppl 3):1–303.
14. Struening EL. Social area analysis as a method of evaluation. In: Struening EL, Guttentag M, eds. *Handbook of Evaluation Research.* Beverly Hills, Calif: Sage; 1975:203–215.
15. McPherson K, Wennberg H, Hovind OB, Clifford P. Small area variations in the use of common surgical procedures: an international comparison of New England, England and Norway. *N Engl J Med.* 1982;307:1310–1314.
16. Roos N. Hysterectomy: variations in rates across small areas and across physicians' practices. *Am J Public Health.* 1984;74: 327–335.
17. Andrews HF, Kerner JF, Zauber A, Mandelblatt J, Pittman J, Struening E. Using census and mortality data to target small areas for breast, colorectal, and cervical cancer screening. *Am J Public Health.* 1994;84:56–61.
18. *1990 Census of Population and Housing.* Washington, DC: US Dept of Commerce, Bureau of the Census; 1993. Publication 1990 CPH-3.
19. Davidson LL, Durkin MS, O'Connor P, Barlow B, Heagarty MC. The epidemiology of severe injuries to children in Northern Manhattan: methods and incidence rates. *Paediatr Perinat Epidemiol.* 1992;6:153–165.
20. *International Classification of Diseases.* 9th revision. Geneva, Switzerland: World Health Organization; 1980.
21. *1980 Census Tracts.* Washington, DC: US Dept of Commerce, Bureau of the Census; 1981. Publication PHC 80-2.
22. Diehr P, Cain K, Connell F, Volinn E. What is too much variation? The null hypothesis in small-area analysis. *Health Serv Res.* 1990;24:741–771.
23. Wennberg J, Gittelsohn A. Small area variations in health care delivery. *Science.* 1973;182:1102–1108.
24. Chassin MR, Brook RH, Park RE, et al. Variations in the use of medical and surgical services by the Medicare population. *N Engl J Med.* 1986;314:285–290.
25. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrica.* 1951;16:297–334.
26. Kerner JF, Andrews H, Zauber A, Struening E. Geographically-based cancer control: methods for targeting and evaluating the impact of interventions in defined populations. *J Clin Epidemiol.* 1988;41:543–553.
27. Cohen J, Cohen P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1983.
28. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods.* Belmont, Calif: Lifetime Learning Publications; 1982.
29. Pless IB, Verreault R, Tenina S. A case–control study of pedestrian and bicycle injuries in childhood. *Am J Public Health.* 1989;79:995–998.
30. Rivera FP, Barber M. Demographic analysis of childhood pedestrian injuries. *Pediatrics.* 1985;76:375–381.
31. Braddock M, Lapidus G, Gregorio D, Kapp M, Banco L. Population, income, and ecological correlates of child pedestrian injury. *Pediatrics.* 1991;88:1242–1247.
32. Greenland S, Maclure M, Schlesselman JJ, Poole C, Morgenstern H. Standardized regression coefficients: a further critique and review of some alternatives. *Epidemiology.* 1991;2:387–392.
33. Morgenstern H. Uses of ecological analysis in epidemiologic research. *Am J Public Health.* 1982;72:1336–1344.
34. Greenland S. Divergent biases in ecologic and individual-level studies. *Stat Med.* 1992;11:1209–1223.
35. Von Korff M, Koepsell T, Curry S, Diehr P. Multi-level analysis in epidemiologic research on health behaviors and outcomes. *Am J Epidemiol.* 1992;135:1077–1082.
36. Grossman DC, Rivera FP. Injury control in childhood. *Pediatr Clin North Am.* 1992;39:471–485.
37. Robinson WS. Ecological correlations and the behavior of individuals. *Am Sociol Rev.* 1950;15:351–357.
38. Davidson LL, Durkin MS, Kuhn L, O'Connor P, Barlow B, Heagarty M. The impact of the Safe Kids/Healthy Neighborhoods Injury Prevention Program in Harlem, 1988 through 1991. *Am J Public Health.* 1994;84:580–586.

Child and Adolescent Drownings in Harris County, Texas, 1983 through 1990

Carla L. Warneke, MS, and Sharon P. Cooper, PhD

ABSTRACT

Objectives. This study described childhood drowning rates and circumstances in Harris County, a large metropolitan area in Texas, and compared case ascertainment between data sources.

Methods. Drowning rates among Harris County residents newborn through 19 years of age were calculated from death certificate data (1983 through 1989), and local childhood drowning hazards were described on the basis of medical examiner data (1983 through 1990). Cases from both sources were compared to determine sensitivity of sources.

Results. The drowning rate among Harris County residents newborn through 19 years of age was 3.8 per 100 000 person-years. The drowning rates among Blacks and Hispanics exceeded that of Whites by 56% and 19%, respectively. The majority of the 196 unintentional drownings occurred in swimming pools. Half of the pool drownings occurred in apartment pools and 33% in private home pools. The medical examiner logbook identified a slightly higher number of drownings than did death certificates. *International Classification of Diseases* external cause of death codes were of limited use in describing drowning circumstances.

Conclusions. Childhood drowning hazards not previously reported were identified, specifically hazards in apartment pools and those among Hispanic children. (*Am J Public Health.* 1994;84:593–598)

Introduction

Drowning is the second leading cause of unintentional injury fatalities among persons newborn through 19 years of age.[1] Drowning patterns vary by geographic area, depending on the local physical and social environment[2]; therefore, community drowning site and victim profiles are necessary to tailor prevention programs appropriate for that community.[3] The three objectives of this study were to determine rates of drowning among Harris County, Texas, residents newborn through 19 years of age, to describe victims and circumstances for child and adolescent drownings occurring in Harris County, and to compare case ascertainment of death certificates and medical examiner files.

Methods

The two data sources used for this study were the Texas Department of Health computerized death certificate file and the Harris County medical examiner files. Data from the death certificates were used to calculate drowning rates for Harris County residents, and data from the medical examiner files were used to describe victims and circumstances for drownings occurring in Harris County. Data from both sources were used to determine the sensitivity of each source for case ascertainment.

Texas computerized death certificate data were available for the years 1983 through 1989. Records for Harris County residents were abstracted from the death certificate file if the person was younger than 20 years of age and the cause of death was drowning. Drowning was defined according to the *International Classification of Diseases* ninth revision (ICD-9) external cause of death codes E830 (accident to water craft causing submersion), E832 (other accidental submersion or drowning in water transport accident), E910 (accidental drowning and submersion), E954 (suicide and self-inflicted injury by submersion), E964 (assault by submersion), and E984 (submersion undetermined whether accidentally or purposefully inflicted).[4]

Injury rates were calculated per 100 000 person-years. The numerators for Harris County rate calculations included only residents of Harris County. Harris County population estimates by age, sex, and race/ethnic group for each of the years 1983 through 1989 were obtained from the Texas Department of Health, Bureau of State Health Data and Policy Analysis.[5] For deaths coded E910 (unintentional drownings excluding boating-related drownings), Harris County resident drowning rates were compared with drowning rates from Texas and the United States. Texas drowning frequencies and annual population estimates were obtained from the Texas Department of Health, Bureau of State Health Data and Policy Analysis,[5] and US drowning frequencies and annual population estimates were obtained from the National Center for Health Statistics.[6] Age- and sex-adjusted rates were calculated with the 1980 US popu-

At the time of the study, both authors were with the School of Public Health, The University of Texas Health Science Center at Houston.

Requests for reprints should be sent to Carla L. Warneke, MS, Texas Children's Hospital, 6621 Fannin St, MC 3-2572, Houston, TX 77030.

This paper was accepted October 14, 1993.

Editor's Note. See related editorials by Pless (p 537) and Kaufer Christoffel (p 539) in this issue.

TABLE 1—Age and Sex-Specific Drowning Rates per 100 000 Person-Years for Harris County, Texas, Residents Newborn through 19 Years of Age, 1983 through 1989, and Male to Female Drowning Rate Ratios (RRs)

	Males			Females			Male: Female RR	Total			
	No. Cases[a]	Rate[b]	95% CI	No. Cases[a]	Rate[b]	95% CI		No. Cases[a]	Population	Rate[b]	95% CI
Age group											
<1 y	6	3.52	1.29, 7.68	17	10.62	6.19, 16.98	0.3	23	330 491	6.96	4.41, 10.44
1–4 y	58	7.98	6.10, 10.41	29	4.17	2.79, 6.01	1.9	87	1 421 914	6.12	4.93, 7.60
5–9 y	24	2.74	1.76, 4.09	8	0.95	0.41, 1.88	2.9	32	1 715 245	1.86	1.28, 2.64
10–14 y	33	4.33	2.98, 6.09	8	1.10	0.47, 2.17	3.9	41	1 488 570	2.75	1.97, 3.73
15–19 y	56	7.48	5.70, 9.79	4	0.54	0.15, 1.39	13.8	60	1 484 415	4.04	3.11, 5.26
Crude rates	177	5.39	4.64, 6.26	66	2.09	1.63, 2.68	2.6	243	6 440 635	3.77	3.32, 4.29
Age-adjusted rates[c]	...	5.49	4.73, 6.38	...	1.91	1.49, 2.45	2.9	...	...	3.74	3.29, 4.24

Note. CI = confidence interval.
[a]Drowning deaths were identified by death certificates coded with the *International Classification of Diseases* (ninth revision) external cause of death code E830, E832, E910, E954, E964, or E984.[4]
[b]Rate per 100 000 person-years; Harris County annual population estimates for 1983 through 1989 were used to calculate rates.[5]
[c]The 1980 US population newborn through 19 years was used as the standard.

lation younger than 20 years of age as the standard. In calculating 95% confidence intervals (CIs), a Poisson distribution was assumed.[7] For rate calculations, race/ethnic group was classified as Anglo (White), Black, or Hispanic. The Texas Department of Health used a computer program to identify Spanish surnames. All Whites with Spanish surnames were classified as Hispanics, and all Whites with non-Spanish surnames were classified as Anglos.

Drownings among persons younger than 20 years were ascertained through the Harris County medical examiner logbook for the years 1983 through 1990. Information concerning victim, site, and circumstance was recorded if the submersion injury *occurred in* Harris County. Drug screens consisted of the urine enzyme-multiplied immunoassay technique (EMIT) screen; positive results were confirmed by ultraviolet spectrometry, mass spectrometry, thin-layer or gas chromatography, or a combination of these. Blood alcohol concentrations were recorded in percentage weight of alcohol per 100 mL of blood. Blood alcohol determinations were considered valid estimates of blood alcohol at time of injury if the individual survived less than 6 hours after submersion and was tested within 24 hours of death.[8]

The computerized death certificate file does not indicate county of submersion injury, but it does list county of death. Therefore, to determine case ascertainment of each data source, drownings among persons newborn through 19 years of age for whom the county of death was Harris County were ascertained from death certificates and from the medical examiner logbooks for the years 1983 through 1989.

The cases from each source were matched on the decedent's date of death, date of birth, age, sex, and residence. Sensitivity was computed for each method of case ascertainment as [cases identified/(cases identified + cases missed) × 100].[9] Death certificate records for cases identified only through the medical examiner logbook were sought by searching the computerized death certificates for a case in which the decedent had the same age, race, sex, residence, date of birth, and date of death as the decedent in the case from the medical examiner file. The medical examiner logs were reexamined in an attempt to locate cases ascertained only by the death certificate file.

Results

Death certificates identified 243 drownings during the years 1983 through 1989 among Harris County residents newborn through 19 years. The county of death was Harris County for 72% of the residents. The remainder died out of state (2%) or in 17 other Texas counties (26%). The crude drowning rate for Harris County residents younger than 20 years was 3.8 per 100 000 person-years (95% CI = 3.3, 4.3), and the age- and sex-adjusted drowning rate was 3.7 per 100 000 person-years (95% CI = 3.3, 4.2).

Drowning rates varied between age groups. Rates were highest among children younger than 5 years (Table 1). The age-adjusted drowning rate for males was 2.9 times higher than the rate for females. The highest drowning rate was for females younger than 1 year. For all other age groups, the rate of drowning for males exceeded the rate for females. Drowning rates varied by race/ethnic group: the age- and sex-adjusted rate per 100 000 person-years for Blacks (5.0; 95% CI = 3.8, 6.4) was 56% higher than that for Whites (3.2; 95% CI = 2.7, 3.9), and the rate for Hispanics (3.8; 95% CI = 2.9, 4.9) was 19% higher than that for Whites. The age-adjusted drowning rate for each sex was highest for Blacks (Table 2).

For drownings coded E910 (unintentional drownings excluding boating-related drownings), age- and sex-adjusted rates per 100 000 person-years in Harris County (3.5; 95% CI = 3.1, 4.0) exceeded the rates for Texas (3.3; 95% CI = 3.1, 3.5) and the nation (2.5; 95% CI = 2.5, 2.6).

For the years 1983 through 1990, the medical examiner logbook identified 214 drownings among persons newborn through 19 years for whom the site of submersion was Harris County. The county of residence in 92% of the cases was Harris County. The nonresidents were from nine other Texas counties (n = 10) and out of state (n = 8). One third of the drowning victims were declared dead at the site of injury.

Files for all cases were retrieved with the exception of one case that had

apparently been misfiled. Limited information available from the medical examiner logbook was used for this case. Deaths were unintentional in 96% (n = 196) of the cases, homicidal in 2% (n = 4), and of undetermined intent in 2% (n = 5). In addition, there were nine newborns who drowned as a result of being delivered into the toilet or bathtub. These birth-related drownings are discussed separately.

Unintentional Drownings

Medical examiner logs identified 196 unintentional drownings. The annual number of unintentional drownings varied from a high of 31 in 1983 to a low of 19 in 1989; however, this variation does not represent a consistent decline in the number of drownings over the 8-year study period. Of the 187 drownings for which the month of injury was known, 81% occurred in the months April through September. Of the 180 cases for which the day of injury was known, drownings occurred most frequently on Sunday (n = 41) and Monday (n = 36).

The age groups with the highest frequency of drownings were children younger than 5 years of age (n = 108) and teenagers 15 through 19 years of age (n = 42). Seventy-four percent of unintentional drowning victims were male. Forty-seven percent of unintentional drowning victims were White, 28% were Black, and 25% were Hispanic. Of the 162 cases in which information on witnesses was reported, 63% of the submersions were not witnessed. The site of submersion was recorded for 192 of the unintentional drownings; 55% occurred in swimming pools or hot tubs, 13% occurred in bathtubs, 3% occurred in buckets or toilets, and 29% occurred in other bodies of water. Drownings in other bodies of water were most frequently witnessed (73%).

Swimming pools and hot tubs. There were 102 drownings in swimming pools and 4 drownings in hot tubs. The children who drowned in hot tubs were aged 16 to 34 months. The children who drowned in pools were most frequently 1- or 2-year-olds.

Swimming pools were the site of the highest percentage of drownings for all age groups under 15 years (Figure 1) and for all race/ethnic groups. Pool drownings occurred in apartment pools (50%), private home pools (33%), public/school pools (11%) and motel/hotel pools (5%). The type of pool in which the drowning occurred varied by race/ethnic group. Drownings among Whites most frequently occurred in private home pools, and drownings among Blacks and Hispanics most frequently occurred in apartment pools.

Together, private home pools and apartment pools represented 91% of pool drownings among preschoolers (children younger than 5 years of age). Among preschoolers, 68% of private home and apartment pool drownings occurred at the home of the decedent, and the remaining 32% occurred at the home of a relative, friend, or neighbor or at a vacant home. Information regarding fencing was seldom recorded and was often incomplete.

Bathtubs. Twenty-five unintentional drownings occurred in the bathtub. Bathtubs were the site of drowning for 33% (n = 17) of the females, compared with 6% (n = 8) of the males. Hispanics (n = 14) drowned in bathtubs at a frequency more than twice that of Whites (n = 6) or Blacks (n = 5). Forty-eight percent of the children who drowned in the bathtub were aged 6 to

TABLE 2—Race- and Sex-Specific Drowning Rates per 100 000 Person-Years for Harris County, Texas, Residents Newborn through 19 Years of Age, 1983 through 1989

	White		Black		Hispanic	
	Females	Males	Females	Males	Females	Males
No. cases[a]	33	83	17	48	16	46
Crude rate[b]	1.97	4.76	2.45	6.87	2.03	5.48
Age-adjusted rate[c]	1.77	4.62	2.39	7.42	1.78	5.76
95% confidence interval	1.22, 2.48	3.70, 5.76	1.39, 3.83	5.47, 9.85	1.02, 2.88	4.21, 7.69

[a]Drowning deaths were identified by death certificates coded with the *International Classification of Diseases* (ninth revision) external cause of death code E830, E832, E910, E954, E964, or E984.[4]
[b]Rate per 100 000 person-years; Harris County annual population estimates for 1983 through 1989 were used to calculate rates.[5]
[c]The 1980 US population newborn through 19 years was used as the standard.

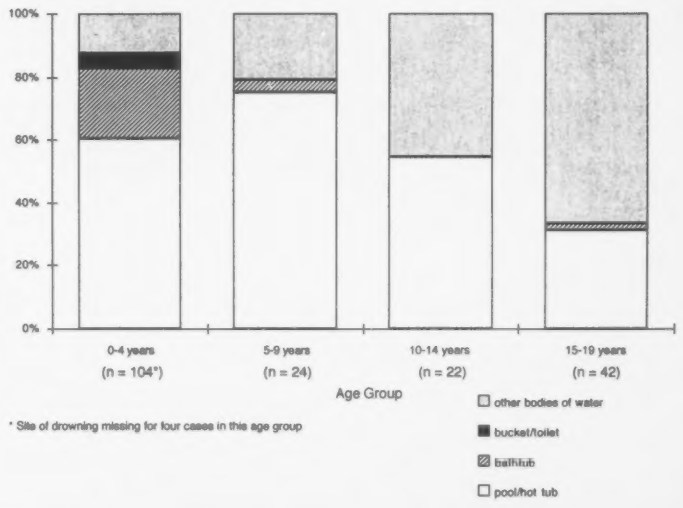


FIGURE 1—Percentage distribution for location of submersion, by age group, for unintentional drownings that, according to medical examiner files, occurred in Harris County, Texas, 1983 through 1990, among persons newborn through 19 years.

TABLE 3—Drugs Detected at Autopsy[a] in Persons Newborn through 19 Years of Age Who Drowned in Harris County, Texas, 1983 through 1990

Drug Category	No. Persons[b]	% of Persons Testing Positive	Drug Detected	
			Frequency	Name
Not likely to be a causal factor in drowning	10	40	2	Acetaminophen (Tylenol)
			3	Caffeine
			1	Ibuprofen
			12	Nicotine
Likely to have been hospital administered	9	36	4	Lidocaine
			3	Phenobarbital
			2	Phenytoin
			1	Thiopental sodium (Pentothal sodium)
Could alter functions needed to prevent injury	6	24	1	Amitriptyline (Elavil)
			1	Cocaine
			1	Codeine
			2	Diazepam (Valium)
			1	Inhalants
			1	Methylene-dioxy-methamphetamine (ecstasy)
			3	Tetrahydrocannabinol (marijuana)

Note. Data were taken from the Harris County medical examiner files.
[a]The medical examiner autopsied or viewed 203 decedents and performed drug screens on 153 decedents. Twenty-five persons tested positive for at least one drug.
[b]Some persons tested positive for more than one drug.

11 months. The two persons older than 4 years of age who drowned in bathtubs had epilepsy or a physical handicap. In 9 cases the child was bathing alone; in 10 cases the child was bathing with someone else, most often one or two preschool-aged children; and information was missing in the remaining 6 cases. Twenty of the children were bathing or playing in the bathtub at the time of injury, three children fell into water left standing in the bathtub, and the activity of two children prior to drowning was not mentioned.

In all cases in which information on supervision was available (n = 23), a lapse in supervision was reported. The length of the lapse was mentioned in 15 cases and ranged from "a few seconds" to 15 minutes, usually 5 minutes or less. The reasons for the lapse were most commonly answering the telephone, performing household chores, or dressing.

Small containers. Four unintentional drownings occurred in buckets, and one drowning occurred in a toilet. All five children were aged 9 to 16 months. The buckets were typically 5-gallon buckets containing water with cleaners used for mopping.

Other bodies of water. Fifty-six unintentional drownings occurred in other bodies of water: rivers (n = 13), bayous (n = 12), lakes (n = 10), ponds or sandpits (n = 6), ditches or irrigation canals (n = 6), creeks (n = 5), and other water sources (n = 4). Eighty-nine percent of these drowning victims were male, and 50% of the drownings occurred to persons aged 15 through 19 years.

Motor vehicle incidents contributed to two drowning deaths, and one drowning resulted from a boating incident. Of the 45 cases in which activity prior to submersion was reported, 42% of the persons were intentionally in the water at the time of injury, 2% drowned during an attempt to rescue another drowning person, 9% were boating or rafting, and 47% did not intend to enter the water. In half of the cases depth of the water (from 5 inches to 25 feet) was mentioned. Information on personal flotation devices was seldom recorded.

Birth-Related Drownings

Medical examiner files identified nine newborn drownings resulting from delivery into the toilet or bathtub. Five of these drownings were unintentional, two were intentional, and two were of undetermined intent. Seven of the nine newborns were Black, one was White, and one was Hispanic. Two of the newborns were premature, and in six cases the mother reported that she did not know she was pregnant before delivery. In three of these cases the mother was living at home with her parents, who also stated that they were unaware of their daughter's pregnancy.

All Drownings

Alcohol and other drugs. Of the 214 cases identified from the medical examiner files, 203 drowning victims were autopsied or viewed, and 84% of these were tested for alcohol, drugs, or both. A blood alcohol analysis was performed on 165 victims, 112 of which met the specified criteria for result representative of blood alcohol concentration at the time of injury. Five of the 112 tested positive, with blood alcohol concentration levels 0.03, 0.04, 0.13, 0.18, and 0.25 g/mL. All positive determinations were for males aged 18 or 19 years. Seventeen other victims tested positive for alcohol; however, in these cases, the representativeness of blood alcohol concentration at time of death either could not be determined or was invalid.

Drug screens were performed on 153 victims, 16% of whom tested positive for at least one drug. When drug screens positive only for drugs that were unlikely to contribute to drowning or drugs that were most likely administered by hospital personnel were eliminated, six persons tested positive for a drug that could have contributed to drowning (one person was also positive for alcohol) (Table 3). With the exception of a newborn who tested positive for diazepam, all persons testing positive for drugs that could have contributed to their deaths were males aged 17 through 19 years. Thirty-five percent (n = 9) of all drownings among 17- through 19-year-old males (n = 26) were drug or alcohol related.

Antecedent medical problems. In 16 cases, antecedent medical problems other than substance use were noted. Three persons experienced acute head injury immediately prior to drowning. One injury resulted from a motor vehicle incident, and two resulted from the individual's hitting his head on the pool while diving or swimming. Eight persons had a history of epilepsy or seizure disorder; four of these were also mentally handicapped. A total of six drownings involved the mentally handicapped. Three drowning victims had other physical handicaps (one was blind, one was

deaf, and one was a quadriplegic as a result of cerebral palsy).

Data Source Comparisons

For the years 1983 through 1989, drowning deaths that occurred in Harris County, regardless of county of submersion, were ascertained from both the medical examiner records and the computerized death certificates. Together, the data sources identified 222 cases. The medical examiner logbook search ascertained 213 cases for a sensitivity of 96%, and the computerized death certificate search ascertained 208 cases for a sensitivity of 94%.

Of the 14 cases identified only by the medical examiner logbook, no death certificate could be identified for 3 (2 homicides and 1 undetermined intent), and 11 were assigned a death certificate ICD-9 code for a cause of death other than drowning: E908.0, weather-related death (n = 3); 345.9, epilepsy (n = 2); E815.0 and E816.0, motor vehicle accident (n = 2); E901.8, accident due to excessive cold (n = 1); 335.0, Werdnig-Hoffman disease, or infantile spinal muscular atrophy (n = 1); 348.1, anoxic brain damage (n = 1); and 348.5, cerebral edema (n = 1). The nine cases identified only by the death certificate file had cause of death codes E830, accident to watercraft causing submersion (n = 1); E910, unintentional drowning (n = 7); and E964, homicide (n = 1).

E-codes were of limited use for determining the circumstances of the drowning. E-codes agreed with the medical examiner data concerning intent of injury in all but one case; however, analysis of E-codes did not describe location or activity of the victim at the time of submersion (Table 4).

TABLE 4—E-Codes Assigned to Death Certificates Compared with Site of Drowning Recorded by Harris County Medical Examiner for Unintentional Drownings of Persons Newborn through 19 Years of Age Who Died in Harris County, Texas, 1983 through 1989

Death Certificate File[a]		Medical Examiner File						
		Boating	Bath	Hot Tub	Pool	Other	Bucket	Total
E830	Accident to watercraft causing submersion	1	...	...	...	...	...	1
E910.4	Accidental drowning or submersion in bathtub	...	1	2	...	...	...	3
E910.8	Other accidental drowning or submersion including drowning in quenching tank or swimming pool[b]	...	26	2	91	50	7	176
E910.9	Unspecified accidental drowning or submersion including fall into water not otherwise specified and drowning not otherwise specified	...	...	...	4	2	...	6
Total		1	27	4	95	52	7	186

Note. Table excludes five cases for which place of injury was unspecified in the medical examiner file.

[a]E-codes are *International Classification of Diseases* (ninth revision) external cause of death codes.[4]

[b]There were no drownings coded E910.0, accidental drowning or submersion while water skiing; E910.1, accidental drowning while engaged in other sport or recreational activity with diving equipment; E910.2, accidental drowning or submersion while engaged in other sport or recreational activity without diving equipment; or E910.3, accidental drowning or submersion while swimming or diving for purposes other than recreation or sport.

Discussion

Among persons younger than 20 years, risk groups for drowning in Harris County, Texas, were similar to those reported in other studies on drowning.[2,10,11] However, a unique finding of this study was that 50% of pool drownings occurred in apartment pools and 33% occurred in private home pools. Most studies have reported that the majority of drownings in pools occurred in private home pools and have not specifically addressed apartment pools.[2,12–14] An exception is a study in King County, Washington, that reported that 22% of the drownings in pools occurred in apartment pools and 37% occurred in private home pools.[3] Further study is needed to determine whether the large percentage of drownings in Harris County apartment pools is explained by the high prevalence of apartment pools in Harris County or by some other factor affecting the safety of the apartment pools.

Drowning prevention is the most effective method for reducing the number of drownings. Although postsubmersion interventions are crucial, one third of the persons who drowned in Harris County never reached the hospital for treatment. Furthermore, survivors of a submersion injury can suffer serious neurologic damage, which can be physically, psychologically, and financially devastating to the child and the child's family.[15]

Individual and societal interventions are two approaches to drowning prevention. Individual interventions include education of the child and the child's supervisor. A national survey of parental knowledge of injury prevention indicated that parents consistently underestimated the risk of drowning to children and that they lacked knowledge about risk reduction beyond "being careful."[16] A survey of pool owners in California found that the proportion of the population endorsing risk reduction measures, such as four-sided fencing of pools, greatly exceeded the proportion who adopted safety measures.[17] This gap between knowledge and practice supports the utilization of social interventions such as legislation for effective pool drowning prevention programs, with health education as a useful adjunct. For each age group and drowning site in the present study, the events that led to drowning were diverse and would require different preventive approaches.

Ascertainment of drowning deaths was similar for both data sources. The medical examiner logs were slightly more sensitive than the death certificate file for identifying drownings, but analysis of medical examiner data, which are not computerized in Harris County, is more time-consuming than analysis of the computerized death certificates. It is possible that both methods of ascertainment missed cases, especially cases in

which persons survived for a period of time to die later of secondary complications. Severely brain-damaged victims have an average life span of 12 to 18 months.[15] If the date of death was more than 1 year from the time of injury, the cause of death should be coded on the death certificate as E929.8, late effects of accidental injury;[4] therefore, such a case would not be included in the death certificate analysis of drownings.

The computerized death certificates were of limited use because of their brevity as well as coding methods. Analysis of the effect of interventions such as mandatory four-sided fencing for pools would not be possible with E-codes, because no specific code exists for pool drownings. Although the medical examiner files were more useful than the computerized death certificates for collecting detailed information on drowning, much of the information needed to evaluate current prevention efforts was missing or included only sporadically. A more systematic data collection method is needed to provide the information necessary to accurately evaluate the effectiveness of established and proposed prevention efforts.

Both data sources ascertained deaths, which represent only a portion of submersion injuries. Studies have estimated that there are from 4 to 18 near-drownings per drowning[3,11,18] and that 6% to 16% of near-drowning survivors suffer from brain damage.[3,18,19] Without information on near-drownings, it is not possible to identify causes of trends in drowning. A decrease in the number of drownings could reflect a real decrease in the number of submersion events, or it may mean that improvements in medical technology or emergency medical services have enabled more of those with submersion injuries to recover intact or with neurologic damage. Surveillance of injury incidence and outcome is needed to monitor current prevention efforts and to assess both individual and societal costs of submersion injury.

Although this study provided potentially useful data concerning victims and circumstances of drowning in Harris County, the study was limited by factors inherent to a descriptive study design. An analytic study design is required to test hypotheses concerning drowning determinants and drowning prevention. Areas needing further investigation are many and include drownings in apartment pools, drownings among Blacks and Hispanics, the effects of swimming lessons on drowning rates, the reduction of alcohol use in aquatic settings among teenaged males, factors influencing adolescents' voluntary use of life vests, the relationship between socioeconomic status and drowning, and the risk of drowning among children with disabilities. In addition, data on the outcome of near-drowning incidents are needed to fully describe the impact of submersion injuries and to evaluate prevention efforts. □

Acknowledgments

The authors acknowledge the cooperation of Harris County Chief Medical Examiner Joseph A. Jachimczyk and staff member Carolyn Anders, who made records available, and the Texas Department of Health, Bureau of Vital Statistics, and Dr William Spears for providing mortality data via the Health Data Library of The University of Texas Health Science Center at Houston School of Public Health. The authors also thank Dr Fred J. Annegers and Dr Sally W. Vernon for their comments and Texas Department of Health epidemiologists David Zane and Dennis Perrotta for their support and advice.

References

1. *National Safety Council Accident Facts.* Chicago, Ill: National Safety Council; 1991.
2. Centers for Disease Control. Childhood injuries in the United States. *Am J Dis Child.* 1990;144:627–646.
3. Quan L, Gore EJ, Wentz K, Allen J, Novack AH. Ten-year study of pediatric drownings and near drownings in King County, Washington: lessons in injury prevention. *Pediatrics.* 1989;83:1035–1040.
4. *International Classification of Diseases, 9th Revision, Clinical Modification.* Ann Arbor, Mich: Commission on Professional and Hospital Activities; 1980.
5. *Texas Vital Statistics.* Austin, Texas: Texas Department of Health; 1983–1989.
6. *Vital Statistics of the United States.* Washington, DC: National Center for Health Statistics; 1983–1988. DHHS publications PHS 87-1101, 87-1122, 88-1101, 88-1122, 90-1101, 91-1101.
7. Haenszel E, Loveland DB, Sirken MG. Lung cancer mortality as related to residence and smoking histories. *J Natl Cancer Inst.* 1962;28:947–1001.
8. Wintemute GJ, Teret SP, Kraus JF, Wright MA. Alcohol and drowning: an analysis of contributing factors and a discussion of criteria for case selection. *Accid Anal Prev.* 1990;22:291–296.
9. Wintemute GJ, Kraus JF, Teret SP, Wright MA. The epidemiology of drowning in adulthood: implications for prevention. *Am J Prev Med.* 1988;4:343–348.
10. Centers for Disease Control. Fatal injuries to children: United States, 1989. *MMWR.* 1990;39:442–451.
11. Wintemute GJ. Childhood drowning and near-drowning in the United States. *Am J Dis Child.* 1990;144:663–669.
12. Press E, Walker J, Crawford I. An interstate drowning study. *Am J Public Health.* 1968;58:2275–2289.
13. O'Carroll PW, Alkron E, Weiss B. Drowning mortality in Los Angeles County, 1976 to 1984. *JAMA.* 1988;260:380–383.
14. Wintemute GJ, Kraus JF, Teret SP, Wright M. Drowning in childhood and adolescence: a population-based study. *Am J Public Health.* 1987;77:830–832.
15. Zamula WW. *Social Costs of Drownings and Near-Drownings from Submersion Accidents Occurring to Children under Five in Residential Swimming Pools.* Washington, DC: US Consumer Product Safety Commission; 1987.
16. Eichelberger MR, Gotschall CS, Feely HB, Harstad P, Bowman LM. Parental attitudes and knowledge of child safety: a national survey. *Am J Dis Child.* 1990;144: 714–720.
17. Wintemute GJ, Wright MA. The attitude-practice gap revisited: risk reduction beliefs and behaviors among owners of residential swimming pools. *Pediatrics.* 1991;88:1168–1170.
18. Pearn JH, Wong RYK, Brown J, Ching YCC, Bart R, Hammar S. Drowning and near-drowning involving children: a five-year total population study from the city and county of Honolulu. *Am J Public Health.* 1979;69:450–454.
19. Centers for Disease Control. Child drownings and near drownings associated with swimming pools—Maricopa County, Arizona, 1988 and 1989. *MMWR.* 1990;39: 441–442.

The Effects of Recall on Estimating Annual Nonfatal Injury Rates for Children and Adolescents

Yossi Harel, PhD, Mary D. Overpeck, DrPH, Diane H. Jones, PhD, Peter C. Scheidt, MD, MPH, Polly E. Bijur, PhD, MPH, Ann C. Trumble, PhD, and John Anderson, PhD

ABSTRACT

Objectives. This study used a recent national population survey on childhood and adolescent nonfatal injuries to investigate the effects of recall bias on estimating annual injury rates. Strategies to adjust for recall bias are recommended.

Methods. The 1988 Child Health Supplement to the National Health Interview Survey collected 12-month recall information on injuries that occurred to a national sample of 17 110 children aged 0 through 17 years. Using information on timing of interviews and reported injuries, estimated annual injury rates were calculated for 12 accumulative recall periods (from 1 to 12 months).

Results. The data show significantly declining rates, from 24.4 per 100 for a 1-month recall period to 14.7 per 100 for a 12-month recall period. The largest declines were found for the 0- through 4-year-old age group and for minor injuries. Rates of injuries that caused a school loss day, a bed day, surgery, or hospitalization showed higher stability throughout recall periods.

Conclusions. Varying recall periods have profound effects on the patterns of childhood injury epidemiology that emerge from the data. Recall periods of between 1 and 3 months are recommended for use in similar survey settings. (*Am J Public Health.* 1994;84:599–605)

Introduction

Injuries have emerged as the principal threat to the health and well-being of children and adolescents. Knowledge of the distribution and determinants of nonfatal injuries in childhood is critical for the development and evaluation of intervention strategies. Population-based surveys are needed to calculate annual injury rates for different subpopulations and/or for different types of injuries. Estimates of nonfatal injury rates in the United States have been based on 2-week recall information obtained from representative samples of households.[1–5] Information ascertained retrospectively through self-reports or proxy reports about exposures or health outcomes may be subject to bias due to differential recall patterns.[6–11] There are two main reasons for recall bias: (1) memory decay—the loss of information due to failure to recall the event, and (2) the "telescoping effect"—the tendency to remember events in the past as if they occurred closer to the present than they really did. Recall bias tends to increase with the enlargement of the recall period inherent in the survey questions.

Since 1957, the National Center for Health Statistics has collected information about injury episodes on the basis of a 2-week recall period.[12] Injuries are identified by asking about reasons for doctor visits or restricted activities within the previous 2 weeks.[3–5] Estimates of annual injury rates are then calculated by multiplying the number of injuries reported for the 2-week recall period by 26 to obtain a 12-month estimate of the numerator. This method may cause the overestimation of injury rates due to telescoping and the magnification of the introduced error by multiplying the number of injuries by 26 to annualize the rate. In addition, the low frequency of injuries in the 2 weeks prior to the interview limits the association of injury with an individual's characteristics. The National Health Interview Survey (NHIS) uses a 2-week recall period for acute conditions because information is lost from failure to remember events over longer periods. Consequently, any attempt to analyze individual or family determinants of injury outcomes based on 2-week recall requires aggregation of data across several years of NHIS surveys to obtain a sample size sufficient for statistical power.

The 1988 Child Health Supplement (CHS) to the NHIS collected information on the occurrence of childhood injuries from an adult respondent, usually the mother. The survey provides the first population-based national data on nonfatal injuries to assist public health

At the time of the study, Yossi Harel was a Visiting Scientist at the Centers for Disease Control and Prevention, Atlanta, Ga, and at the National Institute of Child Health and Human Development, Bethesda, Md. Mary D. Overpeck, Peter C. Scheidt, and Ann Trumble are with the National Institute of Child Health and Human Development, Bethesda, Md. Diane H. Jones and John Anderson are with the Centers for Disease Control and Prevention, Atlanta, Ga. Polly E. Bijur is with the Albert Einstein College of Medicine, Bronx, NY.

Requests for reprints should be sent to Yossi Harel, PhD, Medical Sociology Program, Bar Ilan University, Ramat Gan 52900, Israel.

This paper was accepted September 23, 1993.

Editor's Note. See related editorials by Pless (p 537) and Kaufer Christoffel (p 539) in this issue.

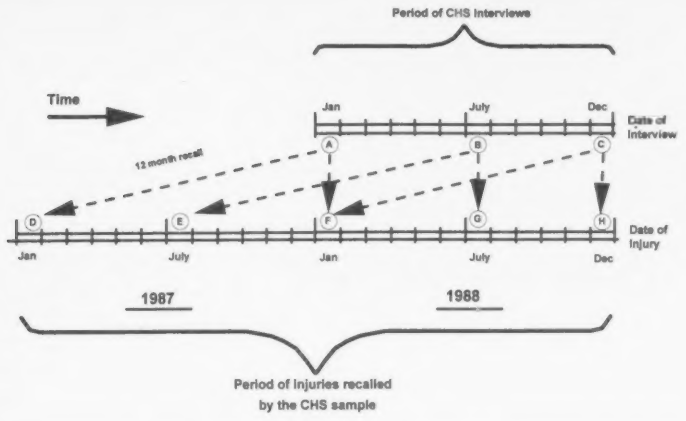


FIGURE 1—Timing of Child Health Supplement interviews and corresponding injury recall periods.

agencies in determining policy objectives and priorities. The NHIS CHS used a 12-month recall period for obtaining information about medically attended injuries rather than the usual 2-week recall period. This longer recall period is likely to produce lower annual estimates than shorter recall periods, owing to memory decay, but may produce higher frequencies of injuries for analysis. The choice of the most appropriate recall period is based on balancing the need for reliable and accurate information with the need for an adequate number of injuries in the data to provide sufficient statistical precision and power.

The purpose of this study was to analyze the effects of using different recall periods on estimated annual injury rates for various subpopulations and for different types of injuries. In so doing, we evaluated the magnitude of the bias and determined the most reliable approach for analysis and interpretation with respect to the recall periods involved.

Methods

Source of Data

The NHIS is a representative sample of the US civilian noninstitutionalized population.[5] The survey collects information about social and economic characteristics, acute and chronic conditions, restricted and limited activity, and physician contacts or any other health care utilization. The 1988 NHIS sample was completed in 47 485 households containing 122 310 persons, 94.9% of whom completed the interviews.[1]

The CHS was added to the core NHIS for the period January 1988 to January 1989 and was designed to produce a representative sample of US children from 0 through 17 years of age. One child in each household with children in the eligible age range was selected, resulting in a sample of 17 110 completed interviews. The information was obtained by face-to-face interview with a responsible adult in the household, usually the child's mother. The survey included such topics as parental history, birth history, school and day care experiences, developmental and behavioral problems, acute and chronic childhood conditions, and other information including a detailed set of items on occurrences of accidents, injuries, and poisonings. Additional information on the methods and design of the 1988 NHIS and CHS surveys is reported elsewhere.[1]

Operational Definition of an Injury

Two screening questions were used: (1) "During the past 12 months, did ____ have an accident, injury, or poisoning that required medical attention?" (2) "How many accidents, injuries, or poisonings did ____ have in the last 12 months that required medical attention?" Beginning with the most recent injury, the respondent was asked to answer questions about the timing and cause of the injury, the place where it happened, the nature of the injury, and restriction of activity, doctor visits, and hospitalizations that resulted from the injury. Information was collected for each of up to 11 injuries reported.

Injuries occurring prior to the recall period that resulted in chronic conditions were not included. Injuries resulting from misadventures or complications of medical care (*International Classification of Diseases* E codes 870–879[13]) were excluded from the analyses. Also excluded were secondary conditions that resulted from injuries already accounted for in a previous injury report.

Timing of the Injury Episode

Interviews were evenly distributed throughout the 1988 calendar year and constituted a nationally representative sample for each individual month of interview. Figure 1 presents the relationship between the date of the interview and the timing of the reported injury. For example, an individual interviewed in January of 1988 (A) may report an injury that occurred within the same week as the interview (F) or an injury that occurred in January 1987, 12 months prior to the interview (D). On the other hand, an individual interviewed 1 year later, in late December of 1988 (C), may report an injury in the same week as the interview (H) or an injury 12 months earlier, in January of 1988 (F). The array of injury events reported for a 12-month recall period covers a calendar period of 2 full years, from January 1987 through December 1988.

To assess the bias resulting from estimates based on different recall periods, the amount of time elapsed between the occurrence of a reported injury and the timing of the interview must be known. Thus, the following timing question was added: "In what month and year did the accident, injury, or poisoning occur?" With information on the month and year in which both the injury and the interview took place, the time interval between the two events can be calculated.

The timing item was added to the CHS instrument only for the last two quarters of the annual sample (i.e., for those interviews occurring between July and December 1988). The respondents interviewed in July 1988 (B) reported recall of injuries occurring in the preceding 12 months, or from July 1987 (E) through July 1988 (G). Similarly, those interviewed in late December 1988 (C)

reported injuries occurring from January 1988 (F) through December 1988 (H). Thus, these data reflect seasonality over a 1.5-year period (from E through H). The present study is based on timing information available for the latter two quarters of the sample. The NHIS sampling technique ensures that the two quarters of available data provide a nationally representative sample.

Calculation of Recall Intervals

The smallest time unit for which timing of the interview and timing of the injury exists is a 1-month period. Therefore, recall time intervals are calculated by subtracting the calendar month of the injury from the interview month. The timing of an interview or the injury event was assumed to be the midpoint of the month in the reporting month. On the basis of this assumption, each calculated interval actually covers a potential recall period of 2 full months.

To illustrate, say an interview that took place in May 1988 reported an injury occurring in April 1988. In this case, the recall interval is calculated as 1 month. However, the sample might include an injury event occurring on April 30 and an interview on May 1, the following day, and a second injury event on April 1 reported from an interview on the last day of May. Despite the fact that the actual time gap between the injury and the interview was only 1 day in the first case and almost 2 months in the second case, both cases represent a 1-month recall period.

The recall period time interval scale involves 12 scores—a score of 1 represents a recall period of 1 day to 2 months, a 2 represents a period of 1 to 3 months, and so on up to a score of 12, which represents a recall period of 11 to 13 months prior to the interview. From a mathematical point of view, this method of calculation creates a smoothing effect between the time-interval scores by virtue of the 1-month overlap between intervals that minimizes random fluctuations occurring between months of interview.

Injuries reported during the calendar month in which the interview took place resulted in a score of 0 months elapsed between the month of the injury and the month of the interview. This interval is incompatible with the other 12 timing categories. First, it covers half the time range covered by other recall intervals (a mean period of only 2 weeks), and second, it results in a range that includes a 12.5-month period. This problem is resolved by adding the injuries with a timing duration of 0 to those reported in the first month. Assuming that both the interviews and the injuries in the 0-month period occurred at the midpoint of the month, the combined 0- and 1-month period is 6 weeks. Therefore, the combined number of reported injuries was multiplied by the proportion of 1.5 months out of a year. Since all subsequent recall periods include the events from the 0-month category in their respective numerators, the multiplier was calculated as follows:

$$12/(j + 0.5),$$

where j = the months elapsed between the injury and the interview. This solution is frequently used in the calculation of life tables. Smith, for example, describes a life-table method in which events occurring in the same month as the interview are excluded from the life tables.[14]

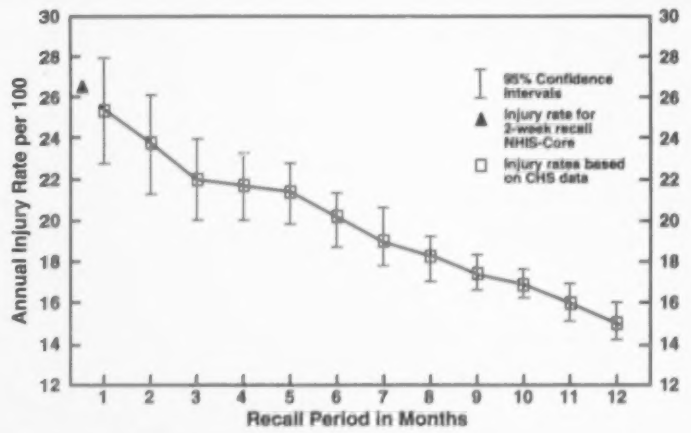


FIGURE 2—Estimated annual injury rates, by recall period, for all children 17 years of age and younger.

Calculating Estimates of Annual Injury Rates per Recall Interval

Weighted annual rates were calculated for recall periods ranging from 1 to 12 months in 1-month increments. Estimated annual injury rates based on a specific recall period are based on the weighted number of injuries reported for that period multiplied by the proportion of the year represented by that recall period and divided by the weighted total number of respondents in the subsample of interest. A constant of $K = 100$ was used to obtain estimated rates per 100 children in the population. Note that the recall periods are mutually inclusive so that a 2-month period, for example, includes the 1-month period. Identical calculations were repeated for different age and sex groups and for different types of injuries. All calculations were carried out using appropriate personal weights. SESUDAAN linear regression methodology was used to calculate 95% confidence intervals for each estimated rate.[15] Approximately 9% of injuries had missing data codes on the timing of the event and thus were not included in the calculation of the numerators for the recall period–specific rates. To maintain the most accurate representation of rates, the data were adjusted for missing data by adding the relative proportion of the injuries with missing data on timing to each recall period–specific rate.

Results

A total of 2854 injuries occurred during the entire survey year. Complete data on 2773 injuries resulted in a total annual injury rate of 16.3 injury episodes per 100 children from 0 through 17 years of age. Of all children reported by their parents to have sustained injuries, 347 (14.9%) had more than one injury. All subsequent results presented here are based on the latter two quarters of the sample for which timing injury information was available.

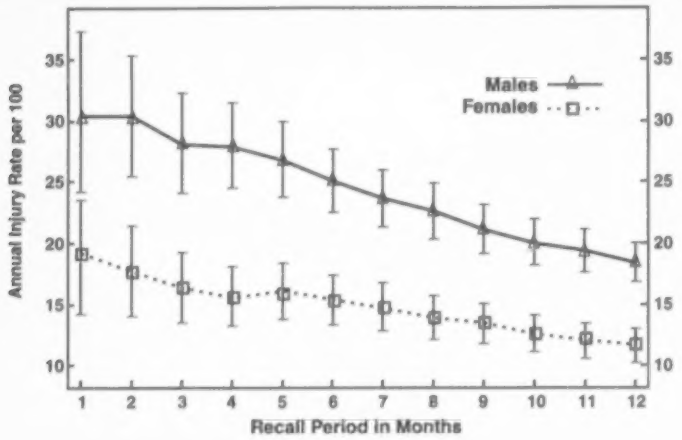


FIGURE 3—Sex-specific estimated annual injury rates, by recall period.

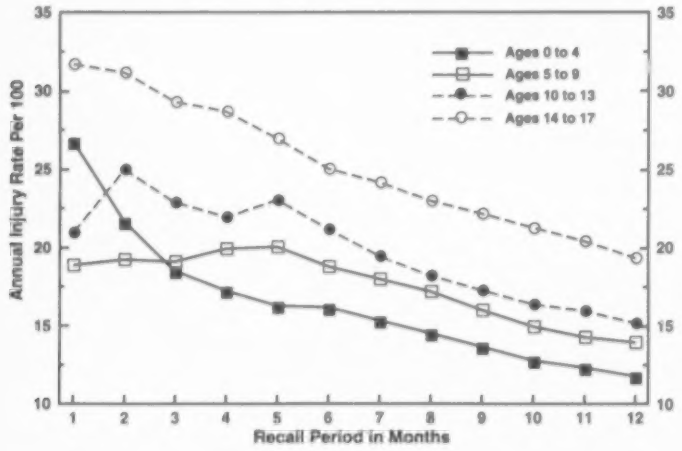


FIGURE 4—Age-specific estimated annual injury rates, by recall period.

Analyses of Recall Bias

Figure 2 shows weighted annual estimated injury rates for all 12 recall periods. These data show a substantial linear decline of the estimated annual rates, from 24.4 per 100 for a 1-month recall period to 14.7 per 100 for a 12-month recall period. The black triangle at top left of the figure represents the 26.4 per 100 injury rate based on a 2-week recall period that most closely approximates the inclusion criteria used in the CHS data (namely, injuries defined as acute conditions, onset during the past 2 weeks, first injury per injury episode, not an adverse effect of medical care, and no doctor visit more than 2 weeks ago). Rates calculated for longer recall periods are based on a larger number of observed injuries in the numerator compared with rates calculated for shorter recall periods. Thus, the confidence intervals are larger for short recall periods and smaller for longer recall periods.

These results demonstrate the critical importance of the length of the recall period in calculating an estimate of injury rates for children. According to these findings, the incidence of injuries among American children can vary from one injury in three children per year, based on a 2-week recall period, to one injury in seven children per year, based on a 12-month recall period.

The results in Figure 2 raise two questions: (1) What are the characteristics of the injuries that tend to be underreported as the recall period is increased? and (2) What are the characteristics of the injuries that tend to be remembered and reported even at 12 months? To answer these questions, we repeated the analyses for different sex and age groups and for injuries of different severity.

Differences in Recall Patterns by Sex

Figure 3 presents estimated annual injury rates per recall period interval by sex. The rates for boys are significantly higher than those for girls for all recall periods. Boys' rates show a sharper decline, from 30 per 100 for a 1-month recall period to about 18 per 100 for a 12-month recall period. A similar slope is shown for girls, whose rates decline from about 17.5 per 100 for a 1-month recall period to about 11 per 100 for a 12-month interval. In both cases, the amount of decline represents about 40% of the rate calculated for the 1-month recall interval.

Differences in Recall Patterns by Age Group

The data presented in Figure 4 show marked differences in recall decline among children in different age groups. Adolescents' rates decline from approximately 32 per 100 for a 1-month recall period to about 19 per 100 for a 12-month recall period. The slope of decline is stable across the 12-month period. The recall of injuries for children aged 5 through 9 and 10 through 13 years is stable for the first 5 months of the recall intervals. However, a remarkable finding is the sharp fall of estimated annual rates for children from 0 through 4 years of age, from a rate of about 27 per 100 for a 1-month recall period to a rate of about 16 per 100 for a 5-month recall period—a fall of over 70% of the decline for the total 12-month period.

Age-specific patterns of recall may differ for boys and girls. To investigate this possibility, we repeated the age-specific analyses for boys and girls separately. The results for age groups

younger than 13 years show similar slopes by sex, despite significant differences in the magnitude of the injury rates between the two sex groups. For adolescents aged 14 through 17 years, the recall of injuries among females appears to be stable across recall intervals, compared with the significant decline found among males. This difference is due, in part, to the difference in the magnitude of rates for adolescents by sex and to the larger number of injuries to be remembered or forgotten by mothers of adolescent boys compared with the smaller number of injuries sustained by girls. Yet the decline for adolescent boys remained sharp and linear across the 12-month period.

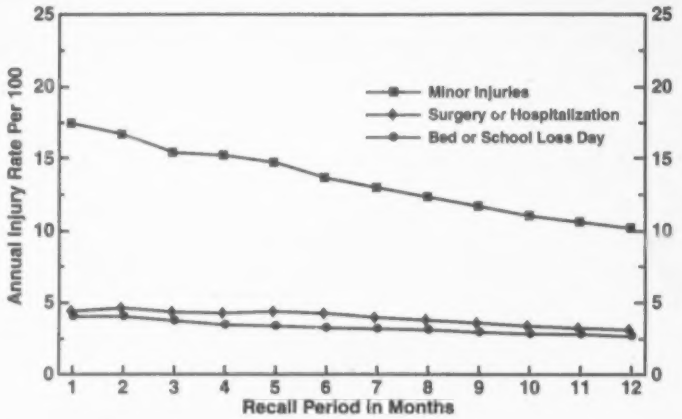


FIGURE 5—Estimated annual injury rates, by severity and recall period.

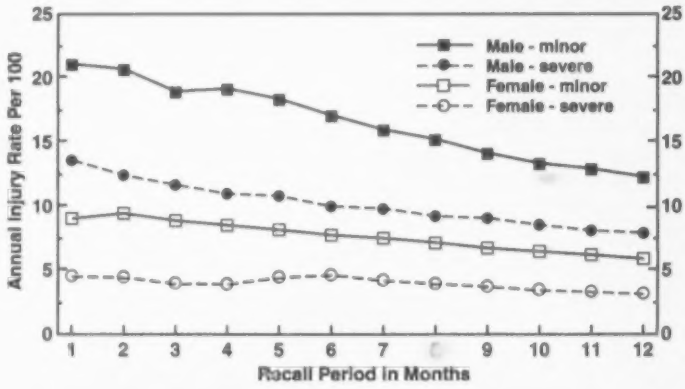


FIGURE 6—Sex-specific estimated annual injury rates, by severity and recall period.

Differences in Recall by Type of Injury Severity Measure

Figure 5 presents estimated annual injury rates by recall intervals for three types of injury severity measures: (1) injuries resulting in surgery or hospitalization (i.e., severity based on medical treatment); (2) injuries resulting in at least 1 full bed day or 1 full school loss day (i.e., severity based on restriction of activity); and (3) injuries that did not result in any of the above consequences (i.e., minor injuries). Note that the first two severity types are not mutually exclusive; however, minor injuries are exclusive of the other two types.

As expected, minor rates decline linearly in magnitude, from 17.5 for a 1-month recall period to about 10 per 100 for a 12-month recall period. On the other hand, patterns are stable for both types of severity of injuries. Injuries that resulted in a doctor visit (not shown in Figure 5) show a slope very similar to that of minor injuries, indicating that reported doctor visits do not explain the variation in recall.

Recall of Age- and Sex-Specific Injury Rates by Injury Severity

To investigate the possibility that the recall differences found for sex and age groups reflect severity differences in the pattern of injuries sustained, we examined sex- and age-specific injury rates for different measures of injury severity by recall periods. Based on the results presented in Figure 5, we constructed a dichotomous measure of injury severity. Injuries resulting in at least 1 bed day, 1 school loss day, surgery, or hospitalization were defined as severe injuries. All other injuries were defined as minor injuries. Figure 6 contrasts the results for minor and severe injuries. Minor injuries have a stronger decline in injury rates across the recall periods for both boys and girls. Minor injury rates for both sexes show similar slopes of decline across the recall periods. However, severe rates appear to be more stable across recall periods for girls than for boys.

Figure 7 shows that severe rates are more stable than minor rates across recall periods for all age groups. This is true in particular for school-aged children between the ages of 5 and 13 years (Figures 7b and 7c). A sharp decline in both minor and severe rates across the first 3 months of recall intervals is found among preschool children. For children from 5 through 9 years of age, a decline in minor injury rates begins only after the 6th month of recall. Minor and severe rates for adolescents decline across recall intervals, although the decline of minor injuries appears to be stronger.

Discussion and Recommendations

The findings presented above suggest that varying recall periods used for calculating national estimates of annual injury rates can be expected to have profound effects on both the magnitude

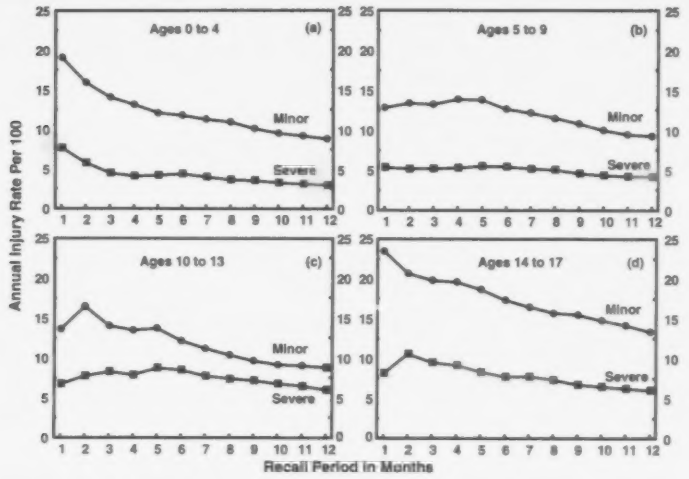


FIGURE 7—Age-specific estimated annual injury rates, by severity and recall period.

and composition of injury estimates. The findings also suggest that recall bias, in varying degrees, is present for all subpopulations, regardless of injury severity, although the effects of recall bias are significantly lower for severe injuries. It is probable that the decline in injury rates across recall periods observed in this study is a function of (1) loss of memory and (2) the "telescoping" tendency to report injury events as if they occurred in more recent calendar months than they actually did. The fall from 26.4 per 100 for a 2-week recall period to 24.4 per 100 for a 1-month recall period suggests that telescoping may be causing overestimation of injury rates by using such a small recall period as 2 weeks. The tendency to recall past events as if they occurred closer to today may cause some respondents to report an event occurring 15 or 16 days ago as if it occurred in the 14 days prior to interview. Reporting error of this type is magnified by a factor of 26 when calculating an annual rate based on a 2-week period. A similar telescoping error results for a 1-month recall period with half the magnitude of bias. Unfortunately, there is no way to test the telescoping hypothesis with the NHIS CHS data alone.

According to the results for all injuries, the incidence of medically attended injuries among American children can vary from a high of one injury in three children per year, based on a 2-week recall period, to a low of one injury in seven children per year, based on a 12-month recall period. By any public health standard, such differences are unacceptable when rates are used to determine priorities in allocating resources for preventive interventions. From an empirical research point of view, such differences imply that results cannot be compared across studies using different recall periods without a clear understanding of the factors affecting the differences in rates.

Sex and Age Differences

Many critical policy decisions are made on the basis of age-specific differences in injury rates. The American Academy of Pediatrics guidelines for well-child supervision,[16] for example, place more emphasis on injury prevention for preschool children than on prevention for older children and adolescents. A recent review of one series of federally funded research projects investigating the patterns and determinants of childhood injuries suggests that nearly 50% of the projects focus on preschool children, compared with only 7% devoted to adolescent injuries.[17] A reason for the emphasis on younger children is suggested from the findings shown in Figure 4. A report based on a 1-month recall period would conclude that children newborn through 4 years experience a relatively high injury rate, second only to the rates of adolescents aged 14 through 17 years. On the other hand, a report basing its findings on a 12-month recall period would conclude that children newborn through 4 years have lower injury rates than children of all other ages.

Severity of Injury

It is assumed that the reporting of severe injuries is less likely to be affected by increased recall periods than is the recall of minor injuries, and that different measures of injury severity will result in different patterns of injury recall, especially when the mother is the source of information. From analyses provided in this report, injuries with serious effects on daily life—for example, loss of school days, surgery, and hospitalization—are less susceptible to memory decay than minor injuries. Consequently, injury patterns emerging from short recall periods have a higher proportion of *minor* injuries than do injury patterns based on longer recall periods.

Recall Bias Issues Not Covered by This Study

The recall bias issues described in this report are applicable only to proxy information obtained from sampled children's parents. It is likely that similar declines in estimated annual injury rates might be found for rates based on responses obtained from the sampled children themselves. However, children, especially adolescents, may or may not recall injuries for reasons different from those of their parents. It is strongly recommended that a similar recall bias analysis be conducted on national data on adolescent injuries based on self-reports. A comparison of recall patterns from such an analysis with those presented above might provide further insight into the different recall forces that affect parents and children.

Implications for Analysis and Interpretation of National Survey Data on Childhood Injuries

The results of this study will help guide the analysis and interpretation of results using information from population-based surveys. In most cases, questions about injuries use one recall period (usually a 12-month period) as a basis for information. Findings based on such data may be influenced by recall bias and should be interpreted accordingly. To

account for recall effects, analyses should be repeated for minor and severe injuries. It is also recommended that researchers conduct sex-specific analyses by the four age categories presented in this paper: preschool children (0 through 4 years), elementary school aged children (5 through 9 years), middle school aged or pubescent children (10 through 13 years), and adolescents (14 through 17 years).*

Implications for the Design of Future Survey Questionnaires

National surveys are a critical source of information for monitoring and studying trends in injuries in childhood and adolescence. The need to include as many injuries as possible in such databases will result in the adoption of a 12-month rather than a 2-week or 1-month recall period. For this reason, it is strongly recommended that retrospective information gathered include the exact time (the date or at least the calendar month) when each reported injury took place. From this timing information an assessment of the magnitude and composition of recall bias can be made as a first step in analyzing the data. Such a procedure will provide more reliable and accurate information, which is essential for developing national intervention priorities. □

*See National Auxiliary Publication Service (NAPS) document 05094 for 5 pages of appendix material (adjustment tables for estimating annual medically attended childhood injury rates). Please order directly from NAPS c/o Microfiche Publications, PO Box 3513, Grand Central Station, New York, NY 10163-3513. Enclose with your order $7.75 for paper copy or $4.00 for microfiche (US funds only, from a US bank). For orders from outside the United States and Canada, add postage of $4.50 for paper copy or $1.50 for microfiche. There is a $15.00 invoicing charge for all orders filled before payment.

Acknowledgments

The authors wish to thank the staff of the National Center for Health Statistics for valuable help during various stages of this project.

References

1. Adams PF, Hardy AM. Current estimates from the National Health Interview Survey: United States, 1988. *Vital Health Stat [10].* 1989;no. 173.
2. The National Committee for Injury Prevention and Control. Injury Prevention: meeting the Challenge. *Am J Prev Med.* 1989;5.
3. Collins, JG. Types of injuries by selected characteristics: United States, 1985–87. *Vital Health Stat [10].* 1990;no. 175.
4. Collins, JG. Impairments due to injuries: United States, 1985–87. *Vital Health Stat [10].* 1991;no. 177.
5. Massey JT. Design and estimation of the National Health Interview Survey, 1985–94. *Vital Health Stat [2].* 1989;no. 110.
6. Cash WS, Moss AJ. Optimum recall period for reporting persons injured in motor vehicle accidents. *Data Eval Methods Res.* 1972;Series 2(50):1–33.
7. Massey JT, Gonzalez JF Jr. Optimum recall periods for estimating accidental injuries in the National Health Interview Survey. *Proc Am Stat Assoc* (Social Statistics Section). 1976:584–588.
8. Funch DP, Marshall JR. Measuring life stress: factors affecting fall-off in the reporting of life events. *J Health Soc Behav.* 1984;25:453–464.
9. Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol.* 1986;123:670–676.
10. Langley JD, Cecchi JC, Williams SM. Recall of injury events by thirteen-year-olds. *Methods Inf Med.* 1989;28:24–26.
11. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol.* 1990;43:87–91.
12. Morgenstern H, Kleinbaum DG, Kupper LL. Measures of disease incidence used in epidemiologic research. *Int J Epidemiol.* 1980;9:97–104.
13. *The International Classification of Diseases.* 9th Revision, Clinical Modification. Hyattsville, Md: National Center for Health Statistics: 1980. DHHS publication PHS 80-1260.
14. Smith DP. Life table analysis. *World Fertil Survey Tech Bull.* 1980;6(1365):6, Table 10.
15. Shah BV. *SESUDAAN: Standard Errors Program for Computing Standardized Rates from Sample Survey Data.* Research Triangle Park, NC: Research Triangle Institute; 1981.
16. *American Academy of Pediatrics Guidelines for Health Supervision.* Elk Grove, Ill: American Academy of Pediatrics; 1985.
17. Scheidt PC. *Childhood Injury Prevention Research: An Internal Report from NICHD.* Bethesda, Md: National Institute of Child Health and Human Development, Center for Research on Mothers and Children, Human Learning and Behavior Branch; 1992.

Characteristics of Adolescent Work Injuries Reported to the Minnesota Department of Labor and Industry

David L. Parker, MD, MPH, William R. Carl, MS, MPH, L. Ronald French, PhD, and Frank B. Martin, PhD

ABSTRACT

Objectives. The purpose of the study was to provide descriptive data and incidence data on adolescent work-related injuries and to determine whether such injuries are underreported to the Minnesota Department of Labor and Industry.

Methods. The study consisted of a 1-year survey of 534 adolescent work-related injuries reported to the Department of Labor and Industry and a cross-sectional survey of 3312 public high school students from throughout Minnesota. The high school survey used an abbreviated questionnaire with a subset of items from the Department of Labor and Industry survey.

Results. Ninety-six percent of the injuries were strains and sprains, cuts and lacerations, burns, bruises and contusions, and fractures. There were 11 hospitalizations; 4 were for burns that occurred during work in restaurants. Eighty workers (15%) reported permanent impairment as a result of their injuries. It was estimated that there were 2268 reportable injuries to working adolescents in Minnesota during the study year.

Conclusions. The most common serious injuries were injuries to the lower back and burns. The demographic characteristics of adolescents whose injuries were reported to the Department of Labor and Industry were similar to those of injured adolescent workers identified through the high school survey. The results suggest that there is substantial underreporting of adolescent work injuries. (*Am J Public Health.* 1994;84:606–611)

Introduction

National estimates of the rate of injury to adolescent workers range from 6.6 per 100 full-time female workers to 12.6 per 100 full-time males.[1] In a recent study from Massachusetts the injury rate for adolescents working part-time was estimated to be equivalent to 16 per 100 full-time workers.[2] Most of these injuries are lacerations, strains and sprains, and heat burns.[1] These data are supported by findings from Minnesota.[3]

Although it has been postulated that many adolescent work injuries may be severe, there are few descriptive data on the severity of adolescent occupational injuries.[4] It has also been suggested that such injuries may be underreported to workers' compensation systems in different states.[5]

In this paper we compare the occupational injury experience of adolescent workers identified through two methods: a survey of public high school students and interviews with adolescents identified via First Reports of Injury compiled by the Minnesota Department of Labor and Industry (DLI). We also evaluate the nature and severity of adolescent work injuries reported to the DLI during the period August 15, 1990, through August 14, 1991. Finally, we analyze the reporting of adolescent work injury events to the DLI, comparing these reports with data obtained from the survey of high school students.

Methods

Department of Labor and Industry Reports

Minnesota law requires that employers file a First Report of Injury for individuals who are fully or partially disabled from work for 3 or more days or who have permanent impairment as the result of a work-related injury or illness. Permanent impairment is defined by Minnesota statute for each condition. Data available from the DLI include personal identifiers for the injured worker, information on the source of the injury (e.g., hot oil, knife), the nature of the injury (e.g., cut, fracture), the event type (e.g., fall, explosion), the part of the body injured, and the date of injury.

For this study, we examined all reports for teenagers younger than 18 years at the time of injury who were injured during the period August 15, 1990, through August 14, 1991. Data from the reports were coded according to standards set by the American National Standards Institute (ANSI).[6] Employer and injury data from the reports were compared with data obtained by telephone interviews.

Parents of injured youths received a letter from the DLI explaining the nature of the Minnesota Department of Health study. A consent form asking for permission to obtain medical records and a postage-paid envelope were included with the letter. When telephone contact was made with the parents, information about family finances and

David L. Parker, William R. Carl, and L. Ronald French are with the Minnesota Department of Health, Minneapolis. Frank B. Martin is with the School of Applied Statistics, University of Minnesota, St. Paul.

Requests for reprints should be sent to David L. Parker, MD, MPH, Minnesota Department of Health, Chronic Disease and Environmental Epidemiology, 717 Delaware St SE, PO Box 9441, Minneapolis, MN 55440-9441.

This paper was accepted June 29, 1993.

Editor's Note. See related editorial by Pless (p 537) in this issue.

health care insurance was obtained. Permission was also obtained to interview the injured youth. If contact was not made with both a parent and the injured worker, a time was arranged during which the interview could be completed. If there was no contact with any member of the family after six calls, a repeat mailing was sent requesting the family to contact the Department of Health.

Information was obtained from injured workers about their work history, the type of work in which they were engaged at the time of injury, the nature and severity of the injury, events surrounding the injury, and the possible impact of the injury on work, school, athletics, and other extracurricular activities. Injuries were coded according to ANSI guidelines as well as the 9th revision of the *International Classification of Diseases*.[7] Up to four types of injury (e.g., laceration, contusion) and four body parts (e.g., finger, elbow) were coded for each injury event. The source of injury was defined as the object, substance, exposure, or bodily motion that directly produced the injury (e.g., floor, hot oil). Event type (e.g., fall, explosion) was defined as the occurrence that led directly to the injury.

After receipt of a medical release form, the Department of Health obtained medical records from individual health care professionals. The nature and severity of the injury, the need for hospitalization, and permanent sequelae were validated by medical record review. Validation was done by an experienced nurse practitioner trained in medical record validation.

Survey of High School Students

In a second study, a sample of 3312 students enrolled in Minnesota public schools was selected through a two-stage stratified cluster sampling procedure.[8] For this sampling procedure, a list of all Minnesota public schools was compiled and grouped into strata according to the number of students in the school. Each stratum contained approximately 25% of the public high school population. Ten schools were contacted from each stratum to obtain administrative approval to conduct the survey. Within each school district, one teacher was chosen to coordinate the survey effort. That individual selected one classroom in each of the school's upper grades (10th, 11th, and 12th) to receive the survey. Data from St. Paul included a sample of 9th-grade students. Survey questions were identical to those used in telephone interviews of workers with injuries reported to the DLI. Again, event type and source of injury were coded according to the ANSI classification system. Analysis was confined to students younger than 18 years at the time they completed the survey.

TABLE 1—Sociodemographics of Working Students in the Minnesota High School Work Injury Survey and Students Filing a First Report of Injury

	High School Survey[a]						First Report of Injury[b]	
	St. Paul		Other Urban[c]		Rural			
	No.	%	No.	%	No.	%	No.	%
Sex								
Female	124	52	365	48	323	47	225	42
Male	116	48	386	52	364	53	309	58
Race								
White	163	68	713	95	665	97	494	92
Asian	34	14	10	1	4	<1	7	1
Other/unknown	43	18	28	4	18	2	33	7
Residence								
Farm	4	2	84	6	503	73	56	10
Non-farm	236	98	667	94	184	27	478	90
Total	240	100	751	100	687	100	534	100

[a]Data represent only students younger than 18 years at the time of the survey. Students in job training and work experience programs are excluded.
[b]Data represent persons younger than 18 years whose work-related injuries were reported to the Minnesota Department of Labor and Industry. Only injuries that occurred between August 15, 1990, and August 14, 1991, were considered in this study.
[c]Urban schools are those in which 13% or fewer students indicated that they lived on a farm.

Analysis

Data were analyzed with SPSS software (SPSS Inc, Chicago, Ill) for Unix operating on a Sun computer. Injuries were considered severe if they limited normal activity and/or resulted in missed work for more than 3 days or resulted in permanent impairment from a herniated lumbar disc, amputation, or other chronic problem such as vision loss, hearing loss, or sensory loss. The possible association of injury severity with type of industry, nature of injury, source of injury, and event type was assessed. Significant differences were assessed by means of a chi-square statistic for two-way contingency tables.[8]

For subjects in the high school survey, total hours worked (exposure) were computed from the date a student began working at his or her most recent job or August 1, 1990, whichever occurred later. Accumulation of work hours ended with the date of survey. A severe injury was defined as one causing the student to miss more than 3 days of normal activity or one that caused a permanent problem. This definition was based on reporting laws in Minnesota that require employers to file a First Report of Injury for individuals who are fully or partially disabled from work for more than 3 days or who have permanent impairment as the result of a work-related injury or illness. Confidence intervals for rate of injury, average hours worked per week, and percent of students working were computed by the jackknife method with data aggregated by classroom.[9]

Estimation of the degree of underreporting of injuries was based on grade- and sex-specific data from the high school survey. These data were used to estimate (1) the percentage of students younger than 18 years in the sample who were working, (2) the percentage who experienced work-related injuries, (3) the percentage of all injuries that were reportable according to Minnesota law, and (4) the total number of reportable injuries for public high school students in Minnesota during the 1990/91 school year.

Results

A total of 742 adolescent work injuries were reported to the DLI from August 15, 1990, through August 14,

TABLE 2—Age at First Job and Hours Worked per Week Reported by Adolescents in Minnesota

	High School Survey[a]				First Report of Injury[b]			
	Females (n = 812)		Males (n = 866)		Females (n = 225)		Males (n = 309)	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Age at first job	14.6	14.4, 14.7	14.1	14.0, 14.3	15.1	13.2, 17.0	14.9	12.7, 17.1
Hours/week in August 1990	25.0	23.6, 26.4	30.1	27.8, 32.5	28.9	26.8, 30.9	32.2	30.5, 34.0
Hours/week in 1990/91 school year	15.3	14.2, 16.3	17.1	16.0, 18.3	20.4	19.1, 21.7	22.6	21.3, 23.8

[a]Data represent only students younger than 18 years at the time of the survey. Students in job training and work experience programs are excluded.
[b]Data represent persons younger than 18 years whose work-related injuries were reported to the Minnesota Department of Labor and Industry. Only injuries that occurred between August 15, 1990, and August 14, 1991, were considered in this study.

TABLE 3—Percentage of Adolescents Performing Specific Types of Work

		High School Survey[a]		First Report of Injury[b]	
Primary Industry	Type of Work	Females (n = 812)	Males (n = 866)	Females (n = 225)	Males (n = 309)
Agriculture	Farm laborer	5	20	1	1
	Yard keeper/landscaper	3	14	0	2
Construction	Laborer	1	8	0	0
Professional	Nurse's aide	5	<1	7	4
Restaurant	Dish buser	11	10	7	6
	Carhop	13	7	8	13
	Cashier	35	16	14	14
	Fast-food cook	9	8	14	12
	Short-order cook	7	11	7	11
	Dishwasher	12	12	11	13
	Waiter/waitress	19	1	5	4
Wholesale/retail	Janitor	9	15	45	43
	Sales clerk	12	5	2	2
	Stock clerk	5	8	15	16
	Bagger/carry-out person	3	8	15	16

Note. Numbers are percentages of column n's
[a]Data represent only students younger than 18 years at the time of the survey.
[b]Data represent persons younger than 18 years whose work-related injuries were reported to the Minnesota Department of Labor and Industry. Only injuries that occurred between August 15, 1990, and August 14, 1991, were considered in this study.
[c]Each participant indicated from one to three activities that best described the work done on the job; therefore percentages may total to more than 100.

1991. For 78 workers (10.5%), neither the worker nor a parent could be reached after repeated attempts by telephone and mail. Of the remaining 664 workers, participation was refused by 47 parents or guardians (7.1%). Contact with 83 parents (12.5%) resulted in a brief interview with the parent only. Therefore, complete survey information was provided by 534 of 664 study participants (80.4%). According to information from the First Reports of Injury, no significant differences in age, sex, nature of injury, or industry of employment were seen between the 208 nonparticipants or incomplete participants and those who completed the entire interview process. These 208 were excluded from further analysis.

Medical records were obtained from 17% of the participants. The only discrepancy between participant interviews and medical records concerned the presence of permanent impairment. The concordance between the reports of study subjects and medical records on permanent sequelae was 63%. Information was not available in the medical record to make a decision about permanent sequelae in 24% of the cases.

Tables 1, 2, and 3 compare the sociodemographic characteristics, work profiles, and job activities of the 534 subjects who had injuries reported to the DLI with those of subjects from the high school survey. Except for subjects from St. Paul, the two samples had similar age, race, and sex distributions for working adolescents. More jobs such as janitor, stock clerk, and bagger occurred in the job descriptions of the DLI sample. School survey respondents from St. Paul were different from all other school survey respondents and from the DLI sample in the following ways: (1) a higher percentage of students were from minority groups and (2) fewer students reported living on farms.

Table 4 shows the injuries most commonly reported to the DLI that resulted in altered activity such as missing school or missing work. In order of frequency of occurrence, these injuries were strains and sprains, cuts and lacerations, burns, bruises and contusions, and fractures. These five groups accounted for 96% of the injuries.

In addition to being the most frequent injury, sprains and strains were the most common cause of severe injury. Injuries to the low and mid back accounted for 73% (147) of all sprain and strain injuries. Most back injuries occurred in janitorial work (48%) and nursing work (21%).

The likelihood of sustaining a back injury was positively associated with the amount of weight individuals lifted at work ($\chi^2 = 17.1$, $P < .0001$). Of the 157 severe injuries, 61 were to the back ($\chi^2 = 14.3$, $P < .001$), and 13 workers reported injuries to the lumbar disc. Larger individuals tended to lift larger amounts of weight than smaller workers. Whether size was determined by body weight ($\chi^2 = 8.1$, $P < .005$) or body mass index[10] ($\chi^2 = 7.9$, $P < .005$), back injuries were more common in smaller workers than larger workers. For those

suffering from severe back injuries, there was a trend for smaller individuals to have more serious lifting injuries than their larger peers. No other injuries were associated with the size of the worker.

Lacerations, cuts, and abrasions were the second most common injuries and most frequently affected the hand (16%) and fingers (61%). Lacerations were often less severe than other injuries. No specific industry or job activity appeared to place workers at increased risk for these types of injury. Although few in number, those working as butchers tended to have severe lacerations.

Burns were the third most common injury. Although only 13% of all injuries were burns, burns contributed to 36% of hospitalizations. Fast-food and full-service restaurants contributed 28 (39%) and 26 (37%) burns, respectively. The most common sources of burn injuries in fast-food restaurants were cooking oils and grills and other cooking equipment. In full-service restaurants, the most common sources of burn injuries were cooking oils and hot water.

Sixty-seven individuals sustained bruises and contusions. These injuries tended to be less severe than others. The most commonly affected sites were the hand and fingers (25%), knee and arm (16% each), and back (15%). The most common event types were being struck by or against objects (37%) and falls (38%). These injuries were positively associated with two jobs: bagger and stock clerk.

Forty-seven individuals sustained fractures, and these injuries were often severe ($\chi^2 = 18.4$, $P < .001$). The most commonly affected site was the finger (43%), followed by the foot and toe (19%). Other significant sites included the skull, face, neck, and back. No specific industry appeared to place individuals at increased risk for fracture injuries.

Eighty workers (15%) indicated that they had permanent impairment as a result of their injuries (Table 5). There were no statistically significant associations between permanent impairments (e.g., amputations) and school days missed. However, the amount of school missed was associated with injury to the back and neck, but not to other body areas. Burns and strains and sprains were the injury types that most commonly resulted in missed school days.

Data from the high school survey were used to calculate the rate of injuries that resulted in adverse outcomes: 16 per 100 000 hours worked for all urban and rural students (95% confidence interval = 12, 19). Although it was not statistically significant, there was a trend for an inverse relationship between risk of injury and grade level. That is, fewer job injuries were observed at higher grade levels.

An estimated 2% of all working students had injuries that resulted in more than 3 days of limited activity and at least some missed work. There were 189 000 public school students in grades 9 through 12 in the 1990/91 school year in Minnesota,[11] and the results of the high school survey indicated that 60% of public high school students (113 400) held a regular job at some time during the study period. Therefore, our results suggest that 2268 reportable injuries occurred to adolescents during the study year ($0.60 \times 189\,000 \times 0.02 = 2268$).

TABLE 4—Nature and Consequences of Injuries to Adolescent Workers Most Commonly Reported to the Minnesota Department of Labor and Industry (DLI)

Result	Strains and Sprains (n = 201)	Cuts and Lacerations (n = 129)	Burns (n = 70)	Bruises and Contusions (n = 67)	Fractures (n = 47)	All Injuries (n = 534)
Limited activity days						
<4	34[a]	65	43	59	19	44[a]
4–13	15	13	27	13	6	15
14–29	20	11	17	18	26	18
30+	31	11	13	10	49	23
Missed school days						
0	65	87	68	68	68	72[a]
<2	11	8	9	15	15	10
2–7	17	5	14	15	11	14
8–90	7	0	9	2	6	4
Missed work days						
0	17	23	19	21	11	17
<4	25	42	23	45	17	30
4–13	15	19	21	12	11	16
14–29	30	11	24	18	23	23
30+	13	5	13	4	38	14

Note. Numbers are percentages of column n's. The five types of injury shown here accounted for 514 (96%) of the 534 injuries for which DLI information was complete. Data for the remaining 20 injuries (not shown) were taken into account in computing the percentages shown for all injuries.
[a]Includes one person for whom injury severity data were incomplete.

TABLE 5—Characteristics of Permanent Impairments for Adolescents with Work Injuries Reported to the Minnesota Department of Labor and Industry

Characteristic	No. of Injuries in Each Category Possessing Characteristic					
	Strains and Sprains	Cuts and Lacerations	Burns	Fractures	Amputations	All Injuries[a]
Chronic pain	13	1	0	3	0	21
Scarring	0	11	31	2	1	45
Limited use of						
Finger(s) or hand(s)	1	6	0	5	3	15
Arm(s)	1	0	0	0	0	2
Leg(s)	1	0	0	1	0	2
Sensory loss	1	6	0	0	1	10

[a]A total of 80 workers reported some type of permanent impairment. The numbers in this column total to more than 80 because some workers reported multiple impairments.

Discussion

This study provided the opportunity to evaluate adolescent work injuries by two methods. The DLI study allowed for an evaluation of the nature and severity of adolescent work injuries, and the school survey provided information on the proportion of adolescents who work and the frequency of adolescent work-related injuries. The DLI study was constrained by probable underreporting and a small number of injuries. The school survey was limited by administrative constraints in obtaining a truly random sample of public high schools. Students from the larger population centers constituted 30% of the school survey population, compared with 41% of the 1990 Minnesota student population.

The primary source of data for the DLI is the First Report of Injury submitted by employers and insurance carriers to the state. In Minnesota, there has been no independent audit of the completeness of workers' compensation reporting by employers. Because the primary reason for reporting is missed work time, use of this measure alone may be misleading if workers return to work prematurely or in a limited capacity.[5,12,13]

In the school survey, disability was measured by the length of time the injury limited a worker's normal activities, the duration of missed work and/or school, and the presence of permanent sequelae. In the DLI study, disability was measured by the duration of missed work and/or school and the presence of permanent sequelae. Each of these methods portrayed a slightly different picture of injury-related disability. Although missed work was the most common limitation experienced by workers whose injuries were reported to the DLI, approximately 12% of the study subjects reported having missed no work.

This study suggests that two thirds of adolescent work injuries were not reported to the Minnesota DLI. This estimate is conservative for several reasons: (1) A large number of adolescent work injuries reported to the DLI did not result in missed work. (2) Agriculture-related work injuries constituted only 1% of all adolescent injuries reported to the DLI, and approximately 40% of male students in rural Minnesota were engaged in agricultural work according to the school survey. (3) There was a 3% high school dropout rate during the 1989/90 school year. Neither these individuals nor individuals who left school in previous years were included in estimates of the number of high school workers or the population upon which this estimate was based. The risk of injury to those who left school is unknown.

Sample size limited our ability to assess the effect of age and stature on injury occurrence and severity. When age was controlled, smaller workers did, however, experience back injuries more frequently and have more severe injuries than larger workers.

Although the epidemiology of low back pain in adolescents is not well understood, it is a matter of concern that approximately 9% of all acute back injuries resulted in herniated lumbar intervertebral discs. A recent prospective survey of 1246 students[14] indicated that 22% of adolescents experienced low back pain within the previous 12 months, and the risk of experiencing low back pain increased with age. This estimate is similar to the 26% reported in a study of secondary school students in Britain.[15] Data from the Minnesota high school survey suggest that 2% of all high school students may experience a work-related back injury each year. This estimate suggests that work contributes to, but is not a primary cause of, back pain in adolescents.

In past studies, burns constituted approximately 11.4% of work-related injuries to adolescents vs 1.8% of all occupational injuries.[3] According to data compiled by the New England Regional Burn Center, for burns requiring hospitalization, 25% of burns to boys and 19% of burns to girls aged 16 through 19 years were work-related. The rate for work-related burns was 143 per 100 000 for boys and 15 per 100 000 for girls.[16] No explanation was presented for the large discrepancy between boys and girls.

In general, work-related burns appear to be a substantial but largely neglected problem in the working population.[16] Existing studies suggest that Black men, food preparers, and younger workers are among those at high risk for work-related burns. Although scalding is the most common cause of industrial burns, flame and electrical burns are the most severe.[17] Grease burns constitute about 8% of all burn injuries in Utah and frequently result in admission to hospital burn units.[18] Deep-fat-fryer oils are maintained at approximately 350°F (177°C), and some filtering systems require as many as four steps of lifting and pouring hot oils.[18] The disposal of oils may require that large quantities of oil be carried in a bucket by hand. No protective equipment is used during either process. Hot oil accounted for 35% of all burn injuries in the DLI study.

Serious work-related injuries were identified in the course of this study. These included numerous cases of herniated intervertebral discs, burns, traumatic brain injuries, and amputations. These data suggest that adolescent work injuries may be a serious problem deserving increased attention. Similarities between the two study samples support underreporting of adolescent work injuries to the DLI. Currently, there is a lack of even the most basic descriptive information on adolescent and child employment in the United States.[4] Owing to the sporadic nature of adolescent work, information on adolescent work and occupational injuries should be obtained by population-based surveys. Estimates based on workers' compensation data leave doubts about the completeness of case ascertainment. □

References

1. Schober SE, Handke JL, Halperin WE, Moll MB, Thun MJ. Work-related injuries to minors. *Am J Ind Med.* 1988;14:585–595.
2. Brooks DR, Davis LK, Gallagher SS. Work-related injuries among Massachusetts children: a study based on emergency department data. *Am J Ind Med.* 1993;24:313–324.
3. Parker DL, Clay RL, Mandel JH, Gunderson P, Salkowicz L. Adolescent occupational injuries in Minnesota. *Minn Med.* 1991;74:25–28.
4. Poliack SH, Landrigan PJ. Child labor in 1990: prevalence and health hazards. *Annu Rev Public Health.* 1990;11:359–375.
5. *Position Papers from the Third National Injury Control Conference.* Atlanta, Ga: Centers for Disease Control; April 1992.
6. National Safety Council and Association of Casualty and Surety Companies. *Method of Recording Basic Facts Relating to the Nature and Occurrence of Work Injuries.* New York, NY: American National Standards Institute; October 4, 1962.
7. *The International Classification of Diseases.* 9th Revision, Clinical Modification. Hyattsville, Md: National Center for Health Statistics; 1980. DHHS publication PHS 80-1260.
8. Conover WJ. *Practical Nonparametric Statistics.* 2nd ed. New York, NY: John Wiley & Sons Inc; 1980:183.
9. Cochran WG. *Sampling Techniques.* 3rd ed. New York, NY: John Wiley & Sons Inc; 1977:174–180.

10. Kahn HS. A major error in nomograms for estimating body mass index. *Am J Clin Nutr.* 1991;54:435–437.
11. *Annual Public School Enrollment Report.* St. Paul, Minn: Minnesota Department of Education; October 1990.
12. National Research Council. *Injury in America: A Continuing Public Health Problem.* Washington, DC: National Academy Press; 1985.
13. *Occupational Safety and Health Statistics, Part II: Supplementary Data System.* Washington, DC: US Dept of Labor, Bureau of Labor Statistics. Handbook of Methods, Bulletin 2134-1.
14. Olsen TL, Anderson RL, Dearwater SR, et al. The epidemiology of low back pain in an adolescent population. *Am J Public Health.* 1992;82:606–609.
15. Fairbank JCT, Pynsent PB, Van Poortvliet JA, Phillips H. Influence of anthropometric factors and joint laxity in the incidence of adolescent back pain. *Spine.* 1984;9:461–464.
16. Rossignol AM, Locke JA, Burke JF. Employment status and the frequency and causes of burn injuries in New England. *J Occup Med.* 1989;31:751–757.
17. Pegg SP, Seawright AA. Burns due to cooking oils—an increasing hazard. *Burns.* 1983;9:362–369.
18. Hayes-Lundy C, Ward S, Saffle JR, Reddy R, Warden GD, Schnebly WA. Grease burns at fast-food restaurants: adolescents at risk. *J Burn Care Rehabil.* 1991;12:203–208.

American Congress of Rehabilitation Medicine to Hold Annual Meeting in June

The American Congress of Rehabilitation Medicine (ACRM) will hold its 71st annual meeting, "Benefits of Rehabilitation—Measure for Measure," June 10 through 12, 1994, at the Hyatt Regency Minneapolis in Minneapolis, Minnesota. The latest theories and techniques in outcome measurement will be featured. The meeting will be an opportunity to showcase what rehabilitation professionals do know about the benefits of rehabilitation and to suggest what still must be demonstrated about the cost-effectiveness of rehabilitation care. The conference is designed for members of ACRM and other healthcare professionals with the responsibilities of delivering high-quality rehabilitation services and achieving positive outcomes for patients.

The education program will feature a variety of presentation methods, including breakfast focus sessions with small-group participation, courses led by experts in specialized areas of study, symposia and scientifically-based panel presentations, workshops providing hands-on experience, and scientific paper and poster presentations. Content will cut across all rehabilitation professions, roles and practice settings.

For more information, contact the ACRM, 5700 Old Orchard Rd, First Floor, Skokie, IL 60077-1057; tel (708) 966-0095; fax (708) 966-9418.

Factors Associated with the Use of Violence among Urban Black Adolescents

Robert H. DuRant, PhD, Chris Cadenhead, BA, Robert A. Pendergrast, MD, MPH, Greg Slavens, BBA, and Charles W. Linder, MD

ABSTRACT

Objectives. The purpose of this study was to examine social and psychological factors associated with the use and nonuse of violence among Black adolescents living in a community with a high level of violent crime.

Methods. Adolescents (n = 225, 44% male) 11 to 19 years of age living in or around nine housing projects in an urban area were administered an anonymous questionnaire.

Results. Self-reported use of violence was associated with exposure to violence and personal victimization, hopelessness, depression, family conflict, previous corporal punishment, purpose in life, self-assessment of the probability of being alive at age 25, and age and was higher among males.

Conclusions. These data support the hypothesis that exposure to violence is associated with adolescents' self-reported use of violence. However, adolescents with a higher sense of purpose in life and less depression were better able to withstand the influence of exposure to violence in the home and in the community. (*Am J Public Health.* 1994;84:612–617)

Introduction

Violence among America's youth has emerged as a significant public health problem[1–7] for which no effective prevention method has been found,[2] and the problem disproportionately involves minority youth.[4] Homicide is the leading cause of death among young Black men 15 to 19 years of age and occurs at a rate nine times that for young White men of the same age.[8] The homicide rate for young Black women has also been found to be 3.5 times higher than that for young White women from the same age group.[3] The National Crime Survey Report estimated that, in 1989, 1.2 million youth 12 to 19 years of age were victims of single-offender crimes and 646 750 were victims of multiple-offender crimes, including robbery, simple assault, aggravated assault, and rape.[9,10] Adolescent violence is a multifaceted problem, and research on its causes is needed to provide requisite data for the development of prevention and intervention programs.

Violence is not a uniquely racial problem,[9–11] and its excess mortality by race is almost entirely accounted for by social factors associated with poverty and unemployment.[2,12–14] Within this social environment, cultural transmission theory proposes that crime and delinquency are learned in interaction with other people, largely within intimate primary groups such as families, peer groups, or gangs.[15] There is extensive evidence that many children and adolescents are continuously exposed to high levels of violence throughout their lives and that this socialization may have a significant effect on the increase in violent behavior observed over the last 40 years.[16] This includes witnessing violence on television and in movies, in the community, and in the home[2,17–21]; the proliferation of gun ownership by adolescents[22–24]; and being a victim of violence, severe corporal punishment, and early physical abuse.[25–40] However, not all adolescents who live in communities with high levels of poverty and unemployment and who are exposed to or are victims of violence engage in violent behavior. Many of these young people demonstrate unexpected resilience in the face of such social insults. Identification of resiliency factors will help in the development of effective prevention programs for urban adolescents who are exposed to social conditions that place them at risk of engaging in violent behavior. The purpose of this study was to examine the relationships between self-reported exposure to or victimization by violence and the use of violence among Black adolescents living in communities characterized by high levels of poverty, unemployment, and violence. Social and psychological factors associated with resiliency among these youth were also examined.

Robert H. DuRant is with the Division of Adolescent/Young Adult Medicine, Children's Hospital, Harvard Medical School, Boston, Mass. Greg Slavens and Charles W. Linder are with the Section of General Pediatrics and Adolescent Medicine, Medical College of Georgia, Augusta. Chris Cadenhead is with the Department of Psychology, Furman University, Greenville, SC. Robert A. Pendergrast is with the Division of Adolescent Medicine, Johns Hopkins University School of Medicine, Baltimore, Md.

Requests for reprints should be sent to Robert H. DuRant, PhD, Division of Adolescent/Young Adult Medicine, Children's Hospital, Harvard Medical School, 300 Longwood Ave, Boston, MA 02115.

This paper was accepted November 10, 1993.

Methods

Subjects

The study was conducted on 225 Black adolescents living in or around nine US Department of Housing and Urban Development housing projects in Augusta, Ga. The subjects ranged in age from 11 to 19 years. A description of the subjects is provided in Tables 1 and 2.

Procedures

The study protocol was approved by the Human Assurance Committee of the Medical College of Georgia. Eight adolescents living in the targeted housing projects were identified by community health workers and enlisted to recruit subjects. The recruiters were asked to contact as many adolescents in their project area as was possible by delivering a parental and youth informed consent form along with information as to when and where the questionnaires were to be administered. The recruiters were paid $1 for each subject who completed a questionnaire. Nonresponse data for those potential subjects who were asked to participate but refused were not collected by the recruiters. To reduce the sampling bias, adolescents living in these housing projects who would not have had the opportunity to attend the scheduled survey sessions because they were participating in summer school (n = 23) or in a housing authority summer enrichment program (n = 40) were invited to participate. More than 90% of eligible subjects (n = 63) in these programs participated in the survey. These three groups of subjects, as well as the recruiters, did not differ on any demographic variable or in self-reported violent behavior. Questionnaires were administered in the housing project community centers, a local school during a summer enrichment program, and an area high school holding summer school. In addition, the questionnaire was administered to a small number of subjects in the conference room of Georgia's Children and Youth Project, which provides health care to children and adolescents living in Department of Housing and Urban Development housing projects. Written parental and subject informed consent was obtained, and the subject's name was not placed on the questionnaire.

To control for variations in reading ability, the questionnaire was read aloud by trained interviewers to groups of subjects ranging in size from 3 to 40; the instrument required between 45 and 60 minutes to administer, depending on the size of the group and the number of questions asked by the subjects. In no cases were subjects' parents present when the questionnaire was administered. Subjects were paid $5 for completing the questionnaire.

TABLE 1—Means, Standard Deviations, and Test–Retest Reliabilities for Demographic Variables

	Mean	SD	Test–Retest Reliability
Age, y	14.4	2.2	1.00
School grade	7.1	2.4	.96
Total number of people in home	5.0	1.9	.98
Head of household's education level	12.5	2.5	.84
Head of household's socioeconomic status	5.8	1.4	.86

TABLE 2—Frequency Distributions of Demographic Variables

	Subjects		Test–Retest Reliability
	No.	%	
Gender			
Male	99	44.0	1.00
Female	126	56.0	
Marital status			
Never married	208	92.4	
Not married but living with partner	16	7.1	1.00
Married	1	0.4	
Employed			
Yes	61	27.4	.84
No	162	72.6	
Religious affiliation			
Yes	156	70.3	.82
No	66	29.7	
Live with parent(s) or relative(s)			
Yes	211	96.3	1.00
No	8	3.7	
Birth order			
Oldest	75	33.3	
Youngest	37	16.4	.89
Middle	100	44.4	
No siblings	13	5.8	
Head of household			
Father, stepfather, grandfather	27	12.0	
Mother, stepmother, grandmother, aunt	191	84.9	1.00
Live on own	3	1.3	
No answer or don't know	4	1.8	

Questionnaire

The questionnaire was constructed from several standardized measures described in the following paragraphs. These instruments were chosen because they have been used extensively among minority youths. In order to assess the stability of the subjects' responses over time, 12 subjects representing each of the survey administration sites were administered the same questionnaire 1 week later. These subjects were given $10 for the retest of the questionnaire. One-week test–retest reliability coefficients were computed on these subjects. For the total sample (n = 225), each constructed scale underwent item analysis and was tested for reliability (internal consistency) with Cronbach's alpha.

Adolescents' use of violence was measured with selected items from the Denver Youth Study Self-Reported Delinquency Questionnaire[41] and items

from the Centers for Disease Control and Prevention's Youth Risk Behavior Survey. Items from these two instruments were combined to create an eight-item scale that was tested for reliability. On the basis of item analysis, seven of the eight items were found to have high internal consistency. These included questions measuring the carrying and use of weapons, fighting and assaults, and involvement in gang fights (see Table 3). The use of violence scale, which ranged from zero to seven, had a Cronbach's alpha of .68 and a 1-week test–retest reliability coefficient of .86 (see Table 4).

Exposure to and victimization by violence in the community were assessed with questions from Richters and Martinez's Survey of Exposure to Community Violence.[33] This survey measures the frequency of exposure to or being a victim of 27 types of violence such as gang violence, selling drugs, burglary, police arrests, assaults, physical threats, sexual assaults, weapon carrying, firearm use, and intentional injuries such as stabbings, shootings, suicides, and murders. On the basis of item analyses, questions measuring exposure to violence had high internal consistency with questions assessing victimization by crime and violence. When separate scales were constructed for being exposed to violence and victimization, the alphas were smaller than for the full scale of 27 items. This 27-item self-report scale, which ranged from 0 (low exposure or victimization) to 116 (high exposure to violence), had a Cronbach's alpha of .85 and a test–retest reliability coefficient of .90.

Exposure to domestic conflict and violence was assessed through a modified version of the Conflict Tactics Scale.[42] This 20-item scale has good concurrent validity when correlated with other measures of parental psychosocial distress.[43] In the present study, 1 item was found to have poor internal consistency with the other 19 items and was discarded. The revised scale ranged from 19 (indicating low domestic conflict) to 71 (representing high conflict). The scale had a Cronbach's alpha of .86 and a test–retest reliability of .97.

The Home Environment Interview, Version II, was used to assess disciplinary activity in the home.[32] This 6-item scale was modified slightly to measure current rather than retrospective patterns of corporal punishment. The scale ranged from 0 to 6 (high disciplinary activity) and had a Cronbach's alpha of .58. However, the test–retest reliability was only .24.

The Children's Depression Inventory was used to measure a variety of symptoms of depression.[44] This 27-item self-report measure has been found to have good reliability and to correlate well with other similar psychological scales. The instrument has a potential range of 0 to 54, with 54 suggesting the maximum level of depression. In this sample, the scale ranged from 0 to 31 and had a Cronbach's alpha of .82 and a test–retest reliability of .84.

The Hopelessness Scale for Children was used to measure subjects' negative expectations about the future.[45] This 16-item scale, which ranges from 0 (indicating low hopelessness) to 12 (indicating a high level of hopelessness), had a Cronbach's alpha of .64 and a test–retest reliability coefficient of .61.

The Purpose in Life test was designed to measure the level of perceived meaning in one's life.[46,47] It has been found to measure a construct distinctly different from locus of control.[48] Each of the 14 items was scored on a 7-point scale so that the minimum ("worst") score was 14 and the maximum was 98. The subjects' scores ranged from 35 to 98, and the scale had a Cronbach's alpha of .86 and a test–retest reliability coefficient of .61.

A 5-point scale was created to determine subjects' certainty of being alive at age 25. The choices ranged from "I am absolutely sure that I will live to be 25 years of age" (scored as 1) to "I am absolutely sure that I won't live to be 25 years of age" (scored as 5). This scale had a test–retest reliability of 1.0.

Subjects' religious activity was assessed with two items. First, the subject was asked whether he or she was a member of any church or religious organization. Second, a 5-point scale was used to determine frequency of church attendance or religious activity. Test–retest reliability was .82 for the former and .60 for the latter.

Four questions were asked to assess family structure. These questions involved the total number of people in the home, whether the subject lived with a parent or adult relative, the subject's birth order, and the identification of the head of the subject's household. All of the questions had a test–retest reliability of .89 or better.

Future aspirations were assessed with two questions. Subjects were asked, "If you could do whatever you wanted, what type of job or profession do you want to have when you grow up?" The answers were scored on a 7-point ordinal scale based on the educational level needed to achieve the professional goal (with 1 being the highest educational level). At a separate point in the questionnaire, the subjects were asked what kind of job they saw themselves doing when they were 25 years of age. This question was scored in the same manner as the preceding question. The test–retest correlation for these questions was .91.

Parents' socioeconomic status was based on the expected educational level needed to qualify the parents for their current job or profession. This was scored on a 7-point ordinal scale, with 1 indicating the highest socioeconomic status level (professional job requiring a doctoral degree), 6 indicating employment not requiring a high school education, and 7 indicating unemployed status.

Statistical Analysis

Interval and ordinal data are summarized as the mean ± standard deviation. Bivariate analyses were conducted with Pearson product moment correlation coefficients, analysis of variance, and chi-square tests. Variables found to be significantly ($P \leq .05$) associated with adolescents' use of violence were analyzed with stepwise multiple regression. Because of missing values on some variables, the regression analysis was conducted on the subjects with complete data (n = 209).

Results

Eighty-four percent of the adolescents reported engaging in at least one form of violent behavior. The percentages of male and female subjects engaging in each type of violence are reported in Table 3. The mean response on the violence scale was 2.1 ± 1.7, with a median score of 2.0 and a skewness of 0.94 ± 0.16. The exposure to violence and victimization scale was normally distributed, ranging from 0 (low) to 116 (high). The mean response of 44.2 ± 22.6 suggests that the subjects in this study had been frequently exposed to and been victims of violence (Table 4).

Male subjects (2.1 ± 1.8) reported engaging in significantly ($P \leq .01$) more violent behavior than female subjects (1.8 ± 1.6). Self-reported use of violence was positively correlated with previous exposure to violence and victimization;

levels of hopelessness, depression, and family conflict; previous corporal punishment; and age. Use of violence was negatively correlated with the purpose in life scale and expectancy of being alive at age 25 (Table 5). Use of violence was not significantly correlated with resiliency factors such as family structure, socioeconomic status, religious behavior, or any other demographic variable. However, a factor indirectly associated with resiliency, higher purpose in life score, was correlated with greater frequency of attendance at religious services ($r = .20$, $P \leq .004$) and a higher socioeconomic status of the head of household ($r = -0.20$, $P \leq .004$). Also, subjects who lived in households in which the head of household was employed had a significantly higher purpose in life score ($P \leq .003$), had fewer feelings of hopelessness ($P \leq .017$), and were more likely to believe they would be alive at age 25 ($P \leq .016$) than adolescents from homes with unemployed heads of households.

When each of the variables found to be significantly associated with use of violence was analyzed with stepwise multiple regression, previous exposure to violence and victimization was the strongest predictor of use of violence, explaining 26.6% of its variation (Table 6). The depression scale was entered as the second variable in the regression model, explaining an additional 3.8% of the variation in use of violence. Gender explained an additional 1.6% of variation in use of violence as the last variable entered into the model. The interaction effects between gender and the other independent variables were not significant. These three variables accounted for 31.1% of the variation in the adolescents' use of violent behavior. Since the other variables significantly correlated with use of violence were also correlated with both exposure to violence and depression, they did not account for any additional variation in the regression model (Table 5). For example, depression was negatively correlated with purpose in life and positively correlated with corporal punishment, family conflict, and hopelessness (Table 5). Exposure to violence and victimization was positively correlated with age, family conflict, depression, and hopelessness and negatively correlated with purpose in life. These data suggest that factors associated with resiliency are not totally independent of the strongest predictor of adolescents' use of violence: their exposure to violence in the past.

TABLE 3—Items Included in the Use of Violence Scale

	Males		Females	
	No.	%	No.	%
Attacked someone they lived with out of anger	44	44.4	66	52.4
Involved in a physical fight in the last 12 months	64	64.6	70	55.5
In the previous 12 months, received an injury during a physical fight requiring medical care	12	12.1	9	7.1
Been involved in a gang fight	25	25.3	15	11.9
Carried a weapon such as a gun, knife, or club in the last 30 days	35	35.4	20	15.8
Ever carried a hidden weapon	41	41.4	31	24.6
Attacked someone with a weapon with the idea of seriously hurting or killing	19	19.2	17	13.5

TABLE 4—Means, Standard Deviations, and Reliability Coefficients for Social and Psychological Scales

	Mean	SD	Test–Retest Reliability	Cronbach's α
Desired socioeconomic status	2.9	1.6	.91	. . .
Anticipated socioeconomic status at age 25	3.5	1.9	.91	. . .
Church attendance	3.5	1.5	.60	. . .
Corporal punishment	0.9	1.3	.24	.58
Conflict Tactics Scale (family conflict)	34.9	9.9	.97	.84
Exposure to and victim of violence	44.2	22.6	.90	.85
Children's Depression Inventory	7.8	6.0	.84	.82
Hopelessness Scale for Children	2.4	2.3	.61	.64
Purpose in Life test	81.2	12.9	.61	.86
Alive at age 25	3.8	0.9	1.00	. . .
Use of violence	2.1	1.7	.86	.68

Discussion

Koop and Lundberg have stated that the solutions to violence are very complex and that research on the causes, prevention, and cures of violence is needed.[1] In this study, we addressed factors associated with Black adolescents' self-reported use and nonuse of violence. As expected, self-reported use of violence by these Black adolescents was significantly correlated with three indicators of previous exposure to violence: self-reported exposure to violence and victimization in the community, degree of family conflict, and severity of corporal punishment and discipline. Because these data were collected in a cross-sectional survey, we cannot imply causation between multiple exposures to violence and adolescents' reported use of violence. These relationships may be covariational in nature, with viewing violence in the community, personally experiencing violence and crime, witnessing violence and conflict among family members in the home, being a victim of severe corporal punishment, and actually using violence dynamically interacting with one another as these events co-occur. These data support the cultural transmission theory that has proposed that adolescents' use of violence is learned within intimate primary groups such as families, peer groups, and other sources for modeling such as gangs.[15] These data also support previous theoretical and empirical work suggesting that experiencing or being a victim of violence will increase the risk that an adolescent will, in turn, use violence against others.[16-21,25-33]

TABLE 5—Pearson Correlation Coefficients (*P* Values) between Black Adolescents' Use of Violence and Social and Psychological Scales

	Exposure to Violence	Age	Alive at Age 25	Purpose in Life	Corporal Punishment	Family Conflict	Depression	Hopelessness
Age	.22 (.001)	...	...	...	...	...	...	...
Alive at age 25	−.11 (NS)	.02 (NS)	...	...	...	...	...	...
Purpose in life	−.18 (.009)	−.04 (NS)	15 (.031)	...	...	...	...	...
Corporal punishment	.12 (NS)	−.10 (NS)	−.10 (NS)	−.26 (.003)	...	...	...	...
Family conflict	.42 (.0001)	−.01 (NS)	−.07 (NS)	−.19 (.005)	.25 (.0001)	...	...	...
Depression	.28 (.0001)	.08 (NS)	−.14 (.005)	−.58 (.0001)	.32 (.0001)	.29 (.0001)	...	...
Hopelessness	.20 (.003)	−.05 (NS)	−.08 (NS)	−.51 (.0001)	.23 (.001)	.33 (.0001)	.55 (.0001)	...
Use of violence	.50 (.0001)	.17 (.01)	−.18 (.007)	−.26 (.0001)	.20 (.004)	.27 (.0001)	.32 (.0001)	.32 (.0001)

Note. NS = not significant.

TABLE 6—Multiple Regression Analysis of Black Adolescents' Use of Violence

	β	R^2 Change	*F*	*P*	*r*
Exposure to and victim of violence	.034	.266	74.99	.0001	.52
Depression	.062	.038	45.12	.0001	.32
Gender	−.443	.016	32.22	.0001	−.15
Constant	.822	...	...	...	...

Multiple R = .566; R^2 = .320; Adjusted R^2 = .311.

Although the depression scale was significantly correlated with the level of violent behavior, we cannot assume that depressed adolescents are more likely to commit violent acts. Because these data were collected by means of a cross-sectional survey, it is also probable that exposure to high levels of violence and involvement in violent activity could result in higher levels of depression among youth. However, these data indicate a definite association between depression and self-reported use of violence independent of the influence of previous exposures to violence.

Although these adolescents were living in a community characterized by high levels of poverty, unemployment, and criminal activity, 15.6% reported that they had never used any form of violence, and an additional 31% reported only one form or type (not event) of previous violent activity. Several variables were found to be associated with resiliency that enabled adolescents to resist engaging in higher levels of violence despite the presence of multiple risk factors. In support of resiliency theory,[34-39] adolescents who reported lower levels of hopelessness, had higher scores on the purpose in life scale, and believed there was a higher likelihood that they would be alive at age 25 were less likely to report engaging in violent behavior. Despite living in an environment that would be expected to facilitate feelings of hopelessness, little chance for successfully changing their life situations, and a chance that they may not live to be adults, many of these adolescents felt the opposite and correspondingly did not engage in violent behaviors. On the basis of research and theoretical work by Jessor,[36] Garmezy,[37] and Werner,[35-39] it would be expected that these relationships would be influenced by family structure, parental socioeconomic status, religious behavior, and parental employment status. None of these variables were associated with the use of violence by adolescents. This may be partly due to the small variation among categories for the family structure, parental socioeconomic status, and parental employment status variables. However, in support of resiliency theory, adolescents who attended religious services more often and whose head of household had a higher socioeconomic status were more likely to have higher purpose in life scores, which, in turn, were associated with less violent behavior. Also, living in a home with an employed head of household was associated with higher purpose in life scores, fewer feelings of hopelessness, and a higher likelihood of believing one would be alive at age 25.

To apply these findings to prevention programs, comprehensive and multifaceted approaches are needed.[4] A lesson that has been learned from the acquired immunodeficiency syndrome epidemic is that cognitive approaches to prevention, such as informing the public about a disease and how to safeguard against infection, do not exert much influence on health-impairing habits among adolescents.[49] The most effective health education programs aimed at preventing and/or changing adolescents' health-impairing behaviors have used a social cognitive approach based on social learning theory.[49,50] This approach assumes that health-impairing behavior is socially learned, purposive, and functional and is the result of the interplay of social-environmental and personal factors. For adolescents to achieve self-directed behavior change, they need to be provided not only with the reasons to alter high-risk behavior but with the means, resources, and social supports to do so. We propose that skill-building violence prevention programs centered on conflict resolution and violence avoidance and aimed at younger adolescents be developed and evaluated and that these programs include strong normative components. However, prevention programs will have a limited impact on adolescent violent behavior if they are not accompanied by institution- and community-level changes in many of the risk factors for engaging in violent behavior (i.e., high levels of violence in the home, community, media, etc.).

For many adolescents coming of age in the 1990s, violence is a daily reality. They experience it in their homes and communities and witness the irresponsible portrayal of violence on prime time television and in movies. Among minority youth living in lower socioeconomic urban areas, the hopelessness of social immobility and a lack of modeling of nonviolent conflict resolution skills in their homes and communities provide scenarios in which there are neither incentives nor skills to avoid the use of violence. At the very least, solving the problem will require both national and personal commitments in initiating prevention programs during childhood and early adolescence. □

References

1. Koop CE, Lundberg GD. Violence in America: a public health emergency. *JAMA.* 1992;267:3076–3077.
2. Roseberg ML, O'Carroll PW, Powell KE. Let's be clear: violence is a public health problem. *JAMA.* 1992;267:3071–3072.
3. Mercy JA, Fenley MA. Forum on youth violence in minority communities: setting the agenda for prevention, summary of the proceedings. *Public Health Rep.* 1991; 106:225–279.
4. Prothrow-Stith D, Weissman M. *Deadly Consequences.* New York, NY: Harper-Collins; 1991.
5. Prothrow-Stith D, Spivak HR. Homicide and violence in youth. In: Wallace HM, Patrick K, Parcel AS, Igol JB, eds. *Principles and Practices of Student Health, Volume 1.* Oakland, Calif: Third Party; 1992:116–123.
6. Mason J, Proctor R. Reducing youth violence—the physician's role. *JAMA.* 1992;267:3003.
7. Novello AC, Shasky J, Froehlke R. A medical response to violence. *JAMA.* 1992;267:3007.
8. Guyer B. An epidemiological overview of violence among children. In: Schwarz DS, ed. *Report of the Twenty-third Ross Roundtable on Critical Approaches to Common Pediatric Problems.* Columbus, Ohio: Ross Laboratories; 1992:3–11.
9. Fingerhut LA, Kleinman JC. International and interstate comparison of homicide among young males. *JAMA.* 1990;263: 3292–3295.
10. Whitaker C, Bastian L. *Teenage Victims: A National Crime Survey Report.* Washington, DC: US Dept of Justice, Bureau of Justice Statistics; 1991. Publication NCJ-128129.
11. Guyer B, Lescohier I, Gallagher SS, Hausman A, Azzara CV. Intentional injuries among children and adolescents in Massachusetts. *N Engl J Med.* 1989;321: 1584–1589.
12. *Homicide Surveillance: High Risk Racial and Ethnic Groups—Blacks and Hispanics, 1970–1983.* Atlanta, Ga: Centers for Disease Control; 1986.
13. Runyan CW, Gerken EA. Epidemiology and prevention of adolescent injury, a review and research agenda. *JAMA.* 1989; 262:2273–2279.
14. Smith DL, DuRant R, Carter TJ. Social integration, victimization and anomie. *Criminology.* 1978;16:395–402.
15. Sutherland EH, Cressey DR. *Criminology.* 10th ed. Philadelphia, Pa: Lippincott; 1978.
16. Centerwall BS. Children, television, and violence. In: Schwarz DS, ed. *Report of the Twenty-third Ross Roundtable on Critical Approaches to Common Pediatric Problems.* Columbus, Ohio: Ross Laboratories; 1992:87–94.
17. Widom CS. Does violence beget violence? A critical review of the literature. *Psychol Bull.* 1989;106:3–28.
18. Gladstein J, Slater EJ. Inner city teenagers' exposure to violence: a prevalence study. *Md Med J.* 1988;37:951–954.
19. Shakoor BH, Chalmers D. Co-victimization of African-American children who witness violence: effects on cognitive, emotional, and behavioral development. *J Natl Med Assoc.* 1991;83:233–238.
20. Martinez P, Richters JE, Benoit M. The NIMH community violence project: II. Children's distress symptoms associated with violence exposure. *Psychiatry.* In press.
21. Widom CS. The cycle of violence. *Science.* 1989;244:160–165.
22. Fingerhut LA, Ingram DD, Feldman JJ. Firearm and nonfirearm homicide among persons 15 through 19 years of age: differences by level of urbanization, United States, 1979 through 1989. *JAMA.* 1992;267:3048–3053.
23. Fingerhut LA, Ingram DD, Feldman JJ. Firearm homicide among Black teenage males in metropolitan counties. *JAMA.* 1992;267:3054–3058.
24. Callahan CM, Rivara FP. Urban high school youth and handguns: a school based survey. *JAMA.* 1992;267:3038–3042.
25. Simcha-Fagan O, Gersten JC. Early precursors and concurrent correlates of illicit drug use in adolescents. *J Drug Issues.* 1986;60:7–28.
26. Bayatpour M, Wells RD, Holford S. Physical and sexual abuse as predictors of substance use and suicide among pregnant teenagers. *J Adolesc Health.* 1992;13: 128–132.
27. Cavaiola AA, Schiff M. Behavioral sequelae of physical and/or sexual abuse in adolescents. *Child Abuse Negl.* 1988;12: 181–188.
28. Riggs S, Alario AJ, McHorney C. Health risk behaviors and attempted suicide in adolescents who report prior maltreatment. *J Pediatr.* 1990;116:815–821.
29. *American Association for Protecting Children: Highlights of Official Child Abuse and Neglect Reporting 1986.* Denver, Colo: American Humane Association; 1987.
30. Powers JL, Eckenrode J. The maltreatment of adolescents. *Child Abuse Negl.* 1988;12:189–199.
31. Bishop V, Woodward K, D'Angelo L. Health risk behavior in urban youth. *J Adolesc Health.* 1992;13:65.
32. Holmes SK, Robins LN. The role of parental disciplinary practices in the development of depression and alcoholism. *Psychiatry.* 1988;51:24–36.
33. Richters JE, Martinez P. *Survey of Exposure to Community Violence, Self-Report Version.* Rockville, Md: National Institute of Mental Health; 1990.
34. Rutter M. Psychosocial resilience and protective mechanisms. *Am J Orthopsychiatry.* 1987;57:316–331.
35. Masters AS, Best KM, Garmezy N. Resilience and development: contributions from the study of children who overcame adversity. *Dev Psychopathology.* 1990;2:425–444.
36. Jessor R. Risk behavior in adolescence: a psychosocial framework for understanding and action. *J Adolesc Health.* 1992;12: 597–605.
37. Garmezy N. Resilience in children's adaptation to negative life events and stressed environments. *Pediatr Ann.* 1991; 20:459–466.
38. Werner EE. High risk children in young adulthood: a longitudinal study from birth to 32 years. *Am J Orthopsychiatry.* 1989;52:72–81.
39. Werner EE. The children of Kauai: resiliency and recovery in adolescence and adulthood. *J Adolesc Health.* 1992;13: 262–268.
40. Kagan J. Etiologies of adolescents at risk. *J Adolesc Health.* 1992;12:591–596.
41. Huizinga D, Esbensen FA, Weiber AW. Are there multiple paths to delinquency? *J Criminal Law Criminology.* 1991;82: 83–118.
42. Straus MA. Measuring intrafamily conflict and violence: the conflict tactics (CT) scales. *J Marriage Fam.* 1979;41:75–88.
43. Wissow LS, Wilson MEH, Roter D, Larson S, Berman H. Family violence and the evaluation of behavioral concerns in a pediatric primary care clinic. *Med Care.* 1992;30:MS150–MS165.
44. Kovacs M. The children's depression inventory (CDI). *Psychopharmacol Bull.* 1985;21:995–998.
45. Kazdin AE, Rodgers A, Colbus D. The hopelessness scale for children: psychometric characteristics and concurrent validity. *J Consult Clin Psychol.* 1986;54: 241–245.
46. Crumbaugh JC, Maholick LT. An experimental study in existentialism: the psychometric approach to Frankl's neogenic neurosis. *J Clin Psychol.* 1964;20:200–207.
47. Crumbaugh JC. The seeking of noetic goals test: a complementary scale to the purpose in life test. *J Clin Psychol.* 1977;33:900–907.
48. Walters LH, Klein AE. A cross-validated investigation of the Crumbaugh purpose-in-life test. *Educ Psychol Meas.* 1980;40: 1065–1071.
49. Bandura A. A social cognitive approach to the exercise of control over AIDS infection. In: DiClemente RJ, ed. *Adolescents and AIDS: A Generation in Jeopardy.* Newbury Park, Calif: Sage; 1992:89–116.
50. Bandura A. *Social Foundations of Thought and Action: A Social Cognitive Theory.* Englewood Cliffs, NJ: Prentice Hall; 1986.

Aggression and Fighting Behavior among African-American Adolescents: Individual and Family Factors

Niki U. Cotten, MPH, Jacqueline Resnick, BA, Dorothy C. Browne, DrPH, Sandra L. Martin, PhD, Donna R. McCarraher, MPH, and Joan Woods, MPH

ABSTRACT

Objectives. This study examined the extent to which individual and family factors are associated with aggression and fighting behavior among African-American middle school adolescents.

Methods. Four hundred thirty-six African-American boys and girls from two middle schools in a predominantly low-income North Carolina school system were surveyed and their school records examined. Information was collected concerning students' aggression levels, school fighting behavior, school suspensions for fighting, attitudes toward violence, perceptions of their families' attitudes toward violence, weapon-carrying behavior, and sociodemographics. Multivariate analyses were employed to predict the students' aggression levels, fighting behavior, and school suspensions.

Results. Factors related to the individual adolescents, such as gender, age, weapon-carrying behavior, and attitudes toward violence, were associated with students' reports of aggression and fighting behavior. Factors related to family and school were associated with school suspension for fighting.

Conclusions. This study suggests that violence prevention programs set in our elementary and middle schools may reduce aggression and fighting among our youth. School teachers and public health practitioners are encouraged to work together in understanding and preventing adolescent violence. (*Am J Public Health.* 1994;84:618–622)

Introduction

In the last decade, rates of violence-related child and adolescent fatalities have increased in the United States. Homicide, the third leading cause of death in adolescents 10 to 19 years of age,[1] accounted for 42% of the deaths among young African-American males during that period.[2] The homicide rate for African-American adolescents is six times that for White adolescents.[3]

Physical fighting is an antecedent behavior for many nonfatal and fatal intentional injuries.[2] Unfortunately, fighting appears to be common among adolescents. A recent survey found that 42% of adolescents reported fighting within the past year, and 26% had carried a weapon during the preceding month.[2]

Much of what is known about violence among African-American youth has been obtained from examining differences between this group and their White counterparts.[4,5] Few studies have examined intragroup differences in aggression and fighting in African-American populations.[6] Therefore, the question addressed in this paper is the extent to which individual characteristics of the adolescent (including gender, age, attitude toward violence, and weapon-carrying behavior) and family factors (including poverty status and students' perceptions of their families' views toward violence) predict aggression and fighting behavior among African-American adolescents.

Methods

Study Setting and Sample

Adolescents attending two middle schools in predominantly low-income African-American neighborhoods in a small North Carolina city were studied. Although only approximately half of the city population is African-American,[7] more than 90% of the children enrolled in the city school system are African-American.

A total of 744 students attended the two schools and therefore were eligible for the study. Ninety-seven percent of these students were African-American (53% male and 47% female). Four hundred forty-seven of these African-American students (60%) received parental consent for their participation in the study, were present in school on the day of the study assessment, and filled out the study questionnaires. Fifty-one percent of these 447 students were male, and 49% were female. Complete information on selected study variables was available for 436 students, of whom 222 (51%) were male and 214 (49%) were female. Therefore, this paper will focus on these 436 students.

Assessment

Student questionnaires. Self-administered questionnaires were given to the students during school, before the initiation of a violence prevention curriculum. Care was taken to establish rapport with

Niki U. Cotten, Jacqueline Resnick, Dorothy C. Browne, and Sandra L. Martin are with the Department of Maternal and Child Health, University of North Carolina at Chapel Hill. Donna R. McCarraher is with Survey Research Associates, Durham, NC. Joan Woods is with the Department of Maternal and Child Health, University of Illinois at Chicago.

Requests for reprints should be sent to Dorothy C. Browne, DrPH, Department of Maternal and Child Health, School of Public Health, CB #7400, University of North Carolina, Chapel Hill, NC 27599-7400.

This paper was accepted December 14, 1993.

the students prior to administration of the questionnaires to maximize the probability of honest responses to sensitive questions. The questionnaires were administered by African-American college students who had been well trained in the administration procedures. Students were assured that their responses would be confidential and that neither teachers nor parents would see their answers. The questionnaire gathered information concerning characteristics of the students (i.e., aggressiveness, fighting behavior, attitudes toward violence, weapon-carrying behavior, gender, and age) and characteristics of the students' families (i.e., students' perceptions of their families' views toward violence).

The questionnaire assessed the students' aggressiveness, using an aggression scale developed on the basis of previous reliable and valid instrumentation.[8–10] The aggression scale listed 12 aggressive behaviors (e.g., threatening peers, destroying property) to which students responded by indicating whether each was never, sometimes, or often true of them. Item responses were summed to create an aggression score, with higher scores indicating more severe aggression (the lowest possible score was 0 and the highest possible score was 24). Internal consistency, as assessed by Cronbach's alpha, was .75, and test–retest reliability over a 3-month period was .60.

The questionnaire also contained a question adapted from the Youth Health Risk Behavior Survey[11] to ask about the students' participation in fights at school. Specifically, students were asked, "While you were at school, has anyone ever attacked or fought with you?"

The students' attitudes toward interpersonal peer violence were assessed with a scale adapted from one developed by the Education and Development Center Inc of Massachusetts. This scale listed 15 statements that expressed either a nonviolent or a violent orientation. Students responded to each item on a 4-point scale, with 1 indicating "disagree a lot" and 4 indicating "agree a lot." Endorsed responses were summed to create a student attitude scale score, with low scores indicating a nonviolent orientation (the lowest possible score being 15) and high scores indicating a violent orientation (the highest possible score being 60). Internal consistency, as assessed by Cronbach's alpha, was .73, and test–retest reliability over a 3-month period was .59.

TABLE 1—Description of the Students

	Female (n = 214)	Male (n = 222)	Total (n = 436)
School, no. (%)			
A	95 (44)	88 (40)	183 (42)
B	119 (56)	134 (60)	253 (58)
Grade, no. (%)			
6	83 (39)	89 (40)	172 (40)
7	53 (25)	70 (32)	123 (28)
8	78 (36)	63 (28)	141 (32)
Poverty, no. (%)			
Yes	106 (50)	118 (53)	224 (51)
No	108 (50)	104 (47)	212 (49)
Fighting at school, no. (%)			
Yes	56 (26)	104 (47)	160 (37)
No	158 (74)	118 (53)	276 (63)
Suspension for fighting, no. (%)			
Yes	31 (14)	46 (21)	77 (18)
No	183 (86)	176 (79)	359 (82)
Carrying weapon to school, no. (%)			
Yes	34 (16)	49 (22)	83 (19)
No	180 (84)	173 (78)	353 (81)
Age, y, mean (SD)	12.5 (1.0)	12.5 (1.0)	12.5 (1.0)
Aggression score, mean (SD)	6.1 (3.5)	6.7 (4.2)	6.4 (3.9)
Student attitude toward violence score, mean (SD)	29.9 (6.5)	32.6 (7.7)	31.3 (7.2)
Perception of family attitude toward violence score, mean (SD)	12.0 (3.4)	12.3 (3.2)	12.2 (5.3)

Student weapon-carrying behavior was assessed with an item from the Youth Health Risk Behavior Survey. Students were asked whether they had ever brought a weapon (e.g., gun, knife, club) to school to protect themselves.

The students' perceptions of their families' views toward violence were assessed on a five-item instrument. Students responded to each item on a 4-point scale, with 1 representing "disagree a lot" and 4 representing "agree a lot." Endorsed responses to items were summed, with low scores indicating that the students believed that the family disapproved of interpersonal student violence (the lowest possible score being 5) and high scores indicating that the students believed that the family condoned interpersonal student violence (the highest possible score being 20). Internal consistency, as assessed by Cronbach's alpha, was .55, and test–retest reliability over a 3-month period was .54.

School record reviews. School records were searched to determine whether the students had been suspended from school for physical fighting, including hitting, biting, and kicking, within the past year.

Families of students were classified as living in poverty if school records showed that students were enrolled in the school free lunch program, indicating that a family of four must have had a monthly income of less than $1452.

Analysis

Multiple linear regression procedures[12] were used to model the students' scores on the aggression scale as a function of school, gender, age, student attitude toward violence scores, student weapon-carrying behavior, students' perceptions of their families' attitudes toward violence scores, and poverty. Regression coefficients and 95% confidence intervals (CIs) were use to measure effect sizes for the predictor variables.

Logistic regression procedures[13] were used to model the students' reports of their school fighting behavior as a function of the predictor variables used in the previously described multiple linear regression model. Odds ratios (ORs) and corresponding 95% confidence intervals were computed. The

TABLE 2—Results of Linear Regression Analysis of Student Aggression Scale Scores

	Regression Coefficient	95% CI
School	−0.27	−0.92, 0.39
Gender	−0.01	−0.67, 0.64
Age	0.50*	0.18, 0.81
Student attitude	0.19*	0.14, 0.24
Weapon carrying	1.10*	0.25, 1.95
Perception of family attitude toward violence	0.07	−0.04, 0.18
Poverty	0.44	−0.21, 1.09

Note. The following variable coding scheme was used in the multiple linear regression model of student report of aggression: aggression (score on aggression scale), school (0 = school A, 1 = school B), gender (0 = male, 1 = female), age (in years), student attitude (score on student attitude toward violence scale), weapon carrying (0 = no, 1 = yes), perception of family attitude toward violence (score on students' perceptions of their families' attitudes toward violence scale), and poverty (0 = no, 1 = yes). CI = confidence interval.
*$P < .05$.

school suspension records for fighting were modeled in a similar manner.

Results

Table 1 presents the characteristics of the sample stratified by gender. Boys constituted 51% of the students. Forty-two percent of the students were from school A, while the remainder were from school B. Forty percent of the students were in sixth grade, 28% were in seventh, and 32% were in eighth. The students ranged in age from 11 to 15 years (mean = 12.5 years). Fifty-one percent of the students' families lived in poverty.

One hundred sixty students (37%) reported that they had been involved in a physical fight at school (26% of the girls, 47% of the boys), and school records showed that 77 (18%) had been suspended from school for fighting (14% of the girls, 21% of the boys). Eighty-three students (19%) reported that they had carried a weapon to school (16% of the girls, 22% of the boys). Overall aggression scores of the students ranged from 0 to 20 (mean 6.4, SD = 3.9), with the mean male scores being higher than the mean female scores. Scores on the student attitude toward violence scale ranged from 15 to 55 (mean = 31.3, SD = 7.2), with the mean male score being higher than the mean female score. The scores on the scale measuring students' perceptions of their families' attitudes toward violence ranged from 5 to 20 (mean = 12.2, SD = 3.3).

Significant positive correlations were found between the students' reports of aggression and (1) their attitudes toward violence ($r = .42$, $P = .0001$) and (2) their perceptions of their families' attitudes toward violence ($r = .22$, $P = .0001$). Students' attitude scores were significantly correlated with their perceptions of their families' attitude scores ($r = .39$, $P = .0001$).

Regression diagnostic procedures detected no collinearity problems among the predictor variables used in the linear regression analysis (tolerance estimates ranged from .79 to .98). The results of the linear regression procedures, presented in Table 2, show that three of the variables focusing on the individual adolescents were significant predictors of aggression. Age was significantly related to aggression; older students were more aggressive than younger students. Student attitude toward violence was also significant; students with violent attitudes reported more aggressive behavior. Finally, students who carried weapons to school scored significantly higher on the aggression scale than those who did not carry weapons to school. The family-related factors and the school factor were not significant predictors of student aggression.

Table 3 shows that when the students' reports of fighting in school were modeled by means of logistic regression procedures, three of the individual factors were significant predictors. Gender was predictive of fighting in school, with girls being less likely to fight than boys (OR = 0.44, 95% CI = 0.29, 0.68). Age was also predictive of student fighting, with older students reporting more fighting than younger students (OR = 1.28, 95% CI = 1.04, 1.57). Student attitude toward violence was predictive of student fighting, with students who had more violent attitudes being more likely to report fighting (OR = 1.06, 95% CI = 1.03, 1.09). The family-related factors and the school factor were not significant predictors of student fighting.

The findings of the logistic model of school records of suspension for fighting, presented in Table 3, show that individual, school, and family factors were all associated with school suspension. Age was related to school suspension; older students were more likely to have been suspended from school for fighting than were younger students (OR = 1.54, 95% CI = 1.19, 1.99). School was significantly related to suspension for fighting, with the students in school B being less likely to be suspended than the students in school A (OR = 0.42, 95% CI = 0.25, 0.70). The family variable of poverty showed a borderline relationship to school suspension, with students of impoverished families being more likely to receive suspensions for fighting in school than students whose families were not impoverished (OR = 1.69, 95% CI = 0.99, 2.85).

Discussion

This study found that individual characteristics of the students, including age, gender, weapon-carrying behavior, and attitude toward violence, were predictive of their reports of aggressive behavior and fighting at school. The findings that age (being an older adolescent) and gender (being male) were associated with student aggression and fighting are consistent with the staggering rates of intentional injury and homicide among African-American adolescent boys.[3] Weapon-carrying behavior was also predictive of student aggression, suggesting that students may gain confidence and security from carrying a weapon and thus may perceive fewer risks to themselves in aggressive situations. Although student attitude toward violence was positively associated with aggression and fighting, students, on average, did not view violent behavior as the most appropriate response to conflict.

These findings suggest that violence prevention programs set in elementary and middle schools may be useful in the prevention of aggression and fighting among youth. Such programs should encourage nonviolent attitudes and teach nonaggressive conflict resolution strategies to give children the tools they need to reduce aggressive behavior. In addition, the findings stress the importance of ensuring that weapons are kept out of classrooms.

Neither family factors nor the school that the students attended were related to student reports of aggression and

fighting. However, a family factor, namely poverty, and school were associated with students having a school suspension record for fighting. Therefore, the question arises as to why students' reports of aggression and fighting were associated with characteristics of the students, while school suspension records for fighting were primarily related to school and family factors.

Discussions with school officials indicated that the discipline policies of the two schools varied greatly, with school A being more likely to use suspension than school B. Therefore, although the students at the two schools did not evidence differential levels of aggressive behavior or fighting at school, differences in school discipline policies resulted in greater proportions of students at one school being suspended for fighting.

The association between poverty and school suspension may be a statistical artifact since the relationship between the two factors is not a strong one. However, if we assume, for the sake of discussion, that this is not the case, other possible explanations arise. One explanation may be that adolescents from impoverished families were more frequently suspended for fighting because, although they fought as often as other adolescents, they were involved in more severe fights that alerted school officials. Post hoc analyses suggested that this was not the case, since 6% of impoverished students who fought reported receiving medical treatment for resultant injuries, as did 6% of nonimpoverished students. An alternative explanation may be that the poorer students were less likely than other students to report aggressive behavior and fighting, thus explaining the discrepancy between self-reported fighting and suspensions for fighting. We do not believe that this is a strong explanation given the great care taken to establish rapport with all students and to assure them of the confidentiality of their answers. Finally, school officials may have been more likely to suspend students from impoverished families than to suspend other students. Future research aimed at providing a better understanding of this finding is encouraged.

Our findings must be viewed with caution given the study limitations. Most of the data were derived from students' self-reports and, therefore, are liable to all of the self-report biases. In addition, it must be stressed that the students' perceptions of their families' attitudes toward violence may not reflect the views the families actually hold. Furthermore, we may not have accurately detected all of the students who were living in impoverished families because of our reliance on enrollment in the free school lunch program as the definition of poverty. Given the social stigma associated with participation in the free lunch program, students eligible for the program may not have enrolled. Therefore, while 51% of the students in this sample were classified as poor because of their free lunch status, the difference in the family financial status between the students whom we classified as impoverished and those we classified as nonimpoverished may not be great. Finally, it should be kept in mind that the analysis strategy that we chose to use (namely, forcing all of the predictor variables simultaneously into the models of the study outcomes) is just one of many approaches that could have been taken. Therefore, different analysis strategies may have led to somewhat different findings.

In conclusion, this study revealed fairly high rates of fighting behavior among African-American middle school students. School teachers and public health practitioners are encouraged to work together to address both the immediate and more subtle factors associated with adolescent violence in order to stem the escalating acts of aggression seen in our schools today. □

TABLE 3—Results of Logistic Regression Analyses of Student Report of Fighting at School and School Suspension Records for Fighting

	Student Report of School Fighting Model		School Suspension for Fighting Model	
	OR	95% CI	OR	95% CI
School	0.76	0.50, 1.15	0.42*	0.25, 0.70
Gender	0.44*	0.29, 0.68	0.69	0.40, 1.18
Age	1.28*	1.04, 1.57	1.54*	1.19, 1.99
Student attitude	1.06*	1.03, 1.09	1.03	0.99, 1.07
Weapon carrying	1.55	0.92, 2.62	1.16	0.61, 2.19
Perception of family attitude toward violence	0.94	0.88, 1.01	0.93	0.85, 1.01
Poverty	0.76	0.50, 1.15	1.69**	0.99, 2.85

Note. The following variable coding scheme was used in the logistic regression model of student report of fighting at school: school fighting (0 = no, 1 = yes), school (0 = school A, 1 = school B), gender (0 = male, 1 = female), age (age in years), student attitude (score on student attitude toward violence scale), weapon carrying (0 = no, 1 = yes), perception of family attitude toward violence (score on students' perception of their families' attitudes toward violence scale), and poverty (0 = no, 1 = yes). The same variable coding was used in the logistic regression model of school suspension record for fighting; however, the outcome of school suspension was coded as 0 = no, 1 = yes. OR = odds ratio; CI = confidence interval.
*$P < .05$; **$P = .05$.

Acknowledgments

This research was partially supported by the Chancellor's Office of the University of North Carolina at Chapel Hill, the Z. Smith Reynolds Foundation, and the University of North Carolina Injury Prevention Research Center.

We would like to thank Dr Betsy Lowman and Mr David Most for their help in data analysis.

References

1. Issac M. *Violence: The Impact of Community Violence on African-American Children and Families.* Arlington, Va: National Center for Education in Maternal and Child Health; 1992.
2. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50213.
3. Centers for Disease Control. Health objectives for the nation: behaviors related to unintentional and intentional injuries among high school students—United States, 1991. *MMWR.* 1992;41:760–772.
4. *Forum on Youth Violence in Minority Communities: Selling the Agenda for Prevention. Summary of the Proceedings.* Atlanta, Ga: Centers for Disease Control; 1991.
5. Wise PH, Kotelchuck M, Wilson ML, Mills M. Racial and socioeconomic disparities in childhood mortality in Boston. *N Engl J Med.* 1985;366:360–366.
6. Rosenburg M, Finley MA. *Violence in America: A Public Health Approach.* New York, NY: Oxford University Press; 1991.

7. *Current Population Reports: Population of the United States.* Washington, DC: US Bureau of the Census; 1990.
8. Rutter M, Tizard J, Whitmore K. *Education, Health, and Behaviour.* London, England: Longman; 1970.
9. Achenbach TM, Edelbrock CS. Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monogr Soc Res Child Dev.* 1981; 46(1).Serial No. 188.
10. Peterson JL, Zill N. Marital disruption, parent-child relationships, and behavioral problems in children. *J Marriage Fam.* 1986;48:295–307.
11. *Youth Health Risk Behavior Survey.* Atlanta, Ga: US Dept of Health and Human Services, Centers for Disease Control; 1990.
12. Kleinbaum DG, Kupper LL, Muller KE. *Applied Regression Analysis and Other Multivariable Methods.* 2nd ed. Boston, Mass: Kent Publishing Co; 1988:124–143.
13. Hosmer B, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons Inc; 1989.

Funding Available for Indoor Air Research

The Center for Indoor Air Research, a sponsor of scientific research on a broad range of indoor air topics, has available its 1994 Request for Applications booklet. Applications for January 1, 1995, funding must be received by *June 1, 1994.*

For more information on the center and/or a copy of the Request for Applications booklet, call or write to the Center for Indoor Air Research, 1099 Winterson Rd, Suite 280, Linthicum, MD 21090; tel (410) 684-3777; fax (410) 684-3729.

Traumatic Child Death and Documented Maltreatment History, Los Angeles

Susan B. Sorenson, PhD, and Julie G. Peterson, BA

ABSTRACT

Objectives. Child abuse is a presumed but largely untested risk factor for child homicide. This research investigated the social and child protective service history of child homicide victims.

Methods. A pairwise matched case–control design was used to assess documented child maltreatment as a risk factor for homicide vs unintentional injury death. Homicide victims aged 0 to 14 years were identified through Los Angeles Police Department case summaries. Control subjects (children who died of an unintentional injury) were matched to case subjects (children who died from homicide) by age, sex, race/ethnicity, and date of death. Case and control subjects were linked with county service records to determine any known history of maltreatment.

Results. A total of 220 children were homicide victims during 1978 through 1987 in the city of Los Angeles. Only one in six children who died (of homicide *or* unintentional injury) or his/her family was known to county social *or* child protective services prior to the death. Recorded history of child protective services was associated with homicide victimization (adjusted odds ratio = 3.40, 95% confidence interval = 1.25, 9.27).

Conclusions. Current service systems need assistance in identifying and protecting children at high risk of homicide. (*Am J Public Health.* 1994;84:623–627)

Introduction

Homicide is the fourth leading cause of death among 1- to 14-year-olds.[1] As Christoffel[2] reported a decade ago, it is the only leading cause of death among children that has increased in rank in the past 30 years,[3] with the exception of human immunodeficiency virus, which appeared recently in child death rankings. Although children make up a small portion of the nation's homicide victims, the years of potential life lost are significant.

Product approaches to the understanding and prevention of intentional injury (e.g., control of firearms)[4–6] are able to address a certain portion of homicides. Previous research indicates, however, that product-oriented countermeasures affect a relatively small portion of the homicides of children because few children, especially those at the youngest ages, are killed with a firearm.[7–15]

A common underlying assumption in prevention programs and in research on child maltreatment is that violence toward children is a continuum ranging from physical punishment through child abuse to homicide.[16–18] The focus of this investigation is whether a history of maltreatment is a risk factor specific to homicide or to child injury death in general.

The present study tested three research questions: (1) What is the social service and child protective service history of homicide victims? (2) Do homicide and unintentional injury deaths of children differ in their history of social service and child protective services? (3) Do the risks represented in the service contact histories of fatally injured children differ according to the homicide victim's relationship to the suspect (i.e., family vs nonfamily)?

Method

Record linkage was used to identify child homicide victims and their history of social and protective service contact. A pairwise matched case–control design was used to assess known child maltreatment history as a risk factor for homicide vs unintentional injury death.

Case and Control Subjects

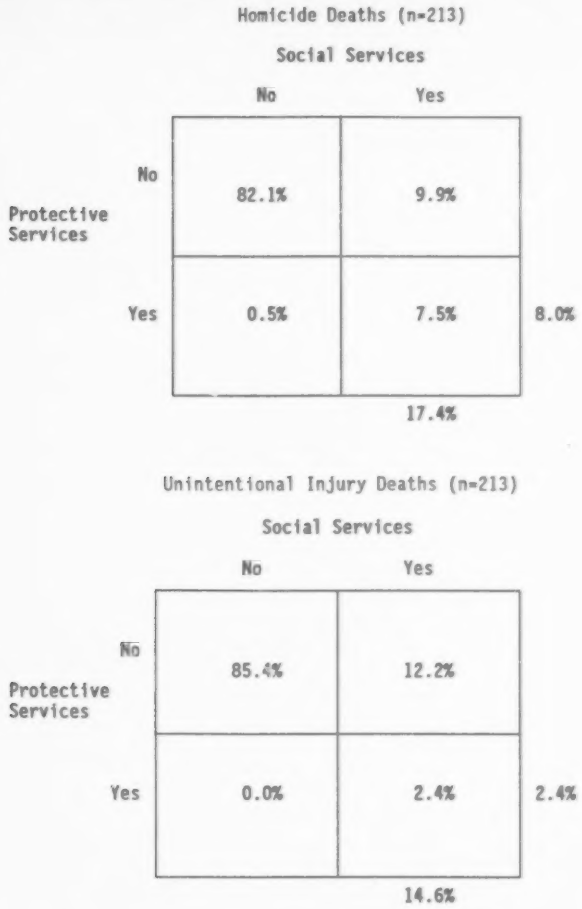
Los Angeles Police Department homicide case summaries for 1978 through 1987 were reviewed; the department is responsible for investigating all deaths that might be classified as homicides in the city of Los Angeles, which has a population base of nearly 3 million. Demographic data and information about the deaths of children from birth to age 14 years were abstracted from the case summaries. A total of 220 cases were determined by the police department to be homicides; these cases make up the homicide sample for the present investigation.

For the matched-control part of the design, California Master Mortality data tapes from 1978 to 1987 (the most recent data available at the time of the study) were used. (Of the deaths labeled homicides by the Los Angeles Police Department, 89.1% were classified as homicides in the California Master Mortality database. Five percent of the total were classified as deaths due to unintentional injuries, 3.2% were classified as deaths due

The authors are with the Southern California Injury Prevention Research Center and the School of Public Health, University of California, Los Angeles.

Requests for reprints should be sent to Susan B. Sorenson, PhD, UCLA School of Public Health, 10833 LeConte Ave, Los Angeles, CA 90024-1772.

This paper was accepted July 7, 1993.



Note. A one-to-one match was made on age, race/ethnicity, sex, and date of death.

FIGURE 1—Social and protective service history of 213 matched pairs of dead Los Angeles children.

to disease, and 2.7% could not be located in the database. Given that no "gold standard" exists, we chose to use the police department determination of homicide.) A Los Angeles child who had died by another sudden, traumatic means—unintentional injury—was matched by age, sex, race/ethnicity, and date of death to each homicide case. Small numbers of deaths at these ages precluded a residential match.

Exact matching was obtained for sex and race/ethnicity (Black, White, Latino, Native American, Indian, Asian). In 73.9% of the time, age matching was exact; in 97.9% of the time, ages were matched within 1 year or less. As for date of death, 97.2% of the matches died within 25 or fewer days of the death of the homicide case. Matches could not be obtained for seven cases: four newborns (Hispanic girl, Hispanic boy, Black girl, White boy), a 6-week-old Asian boy, and two East Indian boys (2 and 5 years old). Of these seven children, one died of exposure, one was pushed from a high place, one was knifed, two died of submersion, and two were beaten to death. When odds ratios are calculated, they are based on a total of 213 cases and an equal number of controls.

Protective and Social Services Contact

County government traditionally assigns the task of protecting vulnerable members of a community (e.g., children and the elderly) to a section of the county governmental structure (often the Department of Public Social Services or the Department of Protective Social Services). In a move to concentrate services for children, Los Angeles County created a separate Department of Children's Services in 1984. This department is responsible for providing and coordinating services for abused (physical and sexual) and neglected children. The Department of Public Social Services was responsible for these tasks prior to 1984 and now attends to the basic social service needs of county residents (including children) by providing food stamps and so forth. Thus, while the organizational structure of Los Angeles County services differs from that of county services in many other areas of the nation, the function of the services remains the same: attending to the needs of maltreated children and providing basic survival needs.

A court order was obtained to access the Department of Children's Services' files. Records were searched to determine whether the department had documented contact with the child or the child's family. When record linkage was established, data regarding the nature of maltreatment and of department services were abstracted. As with the cases, controls were linked with department records to determine the known history of maltreatment. Cases and controls also were linked with Department of Public Social Services records to determine if the child or the child's family had received those services. The nature of services that this department provided was recorded to ascertain whether the contact was to supply basic life necessities or to address maltreatment (as was the department's task prior to 1984). Thus, we ascertained contact with a county agency both for maltreatment/protective services and for basic social services.

Data Analysis

Summary statistics were generated to describe the child homicide victims and the unintentional injury death cases. Conditional logistic regression[19,20] for a one-to-one matched pair design was used to avoid biased parameter estimates by controlling covariates and taking into account the matching variables of age, race/ethnicity, sex, and date of death. Children with no social or protective services history served as the

reference group. Adjusted odds ratios (AORs) and asymptotic 95% confidence intervals (CIs) were calculated for the 213 pairs to measure the difference in risk between those with and those without social service or protective service histories.

Social and maltreatment programs are family-oriented services and might be more effective in preventing homicides of children by family members than in preventing those by nonfamily members. To examine this possibility, we tested whether the children who were killed differed in their history of service contact according to the victim's relationship to the suspect. A chi-square test was used to examine the homogeneity of the adjusted odds ratios obtained for homicides in which a family member was a suspect to those for homicides in which a nonfamily member was a suspect. Because a social definition of family may more accurately describe patterns of relatedness than a biological definition, especially in some racial and ethnic communities,[21] we used a social definition to encompass both biological relatives and primary caretakers (e.g., male companion of mother). Nonfamily suspects included the categories of friend/acquaintance, gang member, drug deal associate, stranger, and unknown. As noted above, the 213 matched children who died of unintentional injury served as the controls.

Results

A total of 220 children were homicide victims in the city of Los Angeles from 1978 through 1987. Modal characteristics of the victims were male sex (64.1%), Black race (49.1%), and age less than 4 years (52.3%). In 91.4% of the cases, at least one suspect was identified, most often a male (73.6%) or a family member (48.2%). The home was the most common site of the homicide (60.0%); the typical victim was shot with a firearm (32.2%) or bludgeoned (31.4%). (A complete descriptive report of child homicides in the city can be found elsewhere.[15])

Among the nonhomicide injury deaths (matched controls), transportation-related injuries were most common (71.4%). Remaining causes of unintentional injury death included burns (8.5%), asphyxiation (7.5%), falls (5.2%), and other miscellaneous means (7.4%).

A significant minority of the families of the murder victims (42.7%) and of the unintentional injury death victims (37.6%) had contact with county services sometime either before or after the child's death. Most of these cases were opened *after* the death of the child. Such action may have served to protect other children in the home but, obviously, could not have had a preventive effect for the dead child. In each type of service or combination of services, a case was more likely to be opened on the family after than before the child's death.

Prior to the death of the child, 17.4% of the homicide victims and 14.6% of the unintentional injury death children had some involvement with the county's social or child protection services. The similarity of these figures suggests that both groups of children came from troubled families. As shown in Figure 1, however, only a small proportion of the homicide cases and even fewer of the controls had a documented history of child protective services (8.0% and 2.4%, respectively).

Of the 17 (8.0%) homicide cases with some record of contact with the county owing to maltreatment prior to their deaths, 16 were also receiving social services assistance. Most (10) of the cases were receiving all of the available social services—cash, MediCal (California's Medicaid program), and food stamps. Another three families were using two of the three services. Of those receiving both child protection and social services, 12 had cases first opened regarding the social services needs. Information about the maltreatment cases (beyond presence/absence and type of record) was limited; only 5 of the 17 cases contained documentation of prior reports. Of the 8 cases that contained any information on siblings, 7 indicated the presence of siblings. After the homicide of their brother or sister, these children typically were removed from the home and placed in foster care.

Adjusted odds ratios (see Table 1) indicated that social needs alone were not related to risk of homicide (vs unintentional injury death). Homicide victims were 3.14 times as likely as unintentional injury victims to have a documented history of social and protective (maltreatment) services before their deaths. Additional analyses indicate that, when maltreatment service history was ignored, a homicide death was not associated with a history of public social services contact (AOR = 1.23; 95% CI = 0.73, 2.07). When general social service history was ignored, however, homicide cases were more likely to have had prior contact with county child protective services (AOR = 3.40; 95% CI = 1.25, 9.27).

TABLE 1—The Risk of Homicide According to Social Service Contact Prior to Death, among 213 Matched Pairs of Dead Los Angeles Children

Service Record[a]	Adjusted Odds Ratio	95% Confidence Interval
Social services only	0.82	0.44, 1.53
Protective services only	. . .[b]	. . .
Protective and social services	3.14	1.14, 8.65

[a]Children with no social or protective services history make up the comparison group.
[b]Odds ratio is undefined because one group had 0%.

One may argue that because county social and protective agencies offer family-oriented services, they might have had a more protective effect on homicides of children by family members. However, after testing the homogeneity of the adjusted odds ratios obtained for homicides in which a family member was a suspect with those for homicides in which a nonfamily member was a suspect, we found no difference by relation of the victim to the suspect.

Discussion

The present study investigates a relatively narrow research question—whether maltreatment is a risk factor specific to homicide or to child injury death in general—among 213 matched pairs of children under 15 years of age who were killed. Many of the children who died of homicide and unintentional injuries (42.7% and 37.6%, respectively) came from families who had contact with county services for child protection or for general social services. However, many of these cases were opened after the death of the child; thus, the death of the child may have alerted social services to the needs of other family members but did not serve as a preventive mechanism for the killed child. Children who died of homicide were 3.14 times as

likely as those whose deaths were unintentional to have had a documented history of both child maltreatment and social service need prior to their deaths.

Mortality risk for abused children was investigated in a recent cohort study that matched (by sex, year of birth, and county of birth) children from birth to age 18 years who were not reported to a Washington State child abuse registry to those who were.[22] Elevated mortality risk following maltreatment was evident for physical abuse and neglect across all categories of death (homicide, transportation injury, other unintentional injury, and disease). Two potential sources of bias are evident in the investigation: selection bias, in that birth certificates could not be located for over half the eligible children (11 846 of the 22 931 children reported to the abuse registry) and possible differential loss to follow-up (without an estimate of geographic mobility, one must assume that there was no differential loss to follow-up between the abused and unabused groups). In other words, the fact that a death certificate was not found does not necessarily mean that the child did not die; rather, it means that the child did not die in the state of Washington. Furthermore, an indicator of the time elapsed between the report of the abuse and the deaths would have provided important information about whether time at reporting is a relevant factor for intervention. Sabotta and Davis[22] conclude that, although intervention by the appropriate government agency does not reduce the risk of death among abused children to that among nonabused children, risk may be even higher without such intervention.

The present investigation adds two pieces of information not available in the Sabotta and Davis study: race/ethnicity of the child and relationship between the child victim and suspect. First, matching on race and ethnicity eliminates the possibility of a spurious association between reports of abuse and homicide; however, the present study shares a limitation with Sabotta and Davis in that socioeconomic information was not available. Second, examining the relationship of the victim to the suspect (i.e., family member vs non–family member) indicates that risk of homicide vs unintentional injury death is not related to the relationship of the victim to the suspect.

Two issues are central in interpreting the finding of no difference between risk of death by a family member and that by a nonfamily member: the nature of the services provided and the maltreatment issues reported to the agency. First, child protection services are family oriented and include the power to remove children from an abusive home. Thus, a documented maltreatment history may not place a child at increased risk for homicide death at the hands of a family member because those children most at risk have been removed from the abusive environment. Additional research is needed to address this important issue. One can note that a history of contact with county protective services does not decrease a child's risk of homicide death to that of a fatal unintentional injury, either. Second, one must recall that child protective services are extended to neglected as well as abused children. Lack of adequate supervision (one aspect of neglect), which, for example, could allow a relatively young child to be outdoors in an area frequented by gangs late at night, may increase a child's risk of homicide by someone outside the family yet still be linked to parental care issues. In this manner, a history of child maltreatment that had come to the attention of county services could be linked to an increased risk of child homicide (compared with a fatal unintentional injury) at the hands of a nonfamily member.

Study Limitations and Suggestions for Future Research

Although documented maltreatment was found to be associated with increased risk of homicide in this investigation, causal factors were not ascertained. Documented maltreatment may serve as a proxy for other factors, including but not limited to socioeconomic status, substance use by caretakers, and family disruption and discord.

The difficulty of selecting appropriate controls is well documented, and we acknowledge that some of the matched controls actually may have been unrecognized homicide victims. For example, a portion of the asphyxiation deaths that were labeled unintentional may have been intentional strangulations. (Child death review teams may enhance consensus on which cases of child death are labeled homicide.[23–25]) Such potential misclassification error would reduce the likelihood of detecting differences between the groups because deaths that actually were homicides would be included in the controls; thus, misclassifying homicides as unintentional injury deaths would underestimate the effect.

Under ideal circumstances, which would include greater resources for research and unlimited access to service delivery systems, several control groups could be used in an effort to detect additional differences (see, e.g., the work of Haddon and colleagues[26]). Depending on the research question, additional control groups could be made up of children who died of noninjury causes, living children with nonfatal intentional injuries, living children with nonfatal unintentional injuries, uninjured living children, and other groups. Investigators must be sensitive, however, to the strain that research activities place on an already overburdened service delivery system.

The present research has limited applicability to the homicide of infants. Matches could not be obtained for any of the newborns, who accounted for four of the seven cases that could not be matched. Risk factors for the homicide of newborns probably follow patterns different from those for the homicide of older children.[27,28]

Implications

Children are a unique segment of our population in that specialized services such as foster care and child abuse prevention agencies have been developed for and are available only to them. Secondary prevention services (e.g., home visitor programs) targeting services to specific high-risk groups have been shown to maintain the family and to reduce abusive behaviors toward children.[29–31] However, identifying family structure variables associated with child homicide remains a complex task.[32] High-risk groups must be identified in sufficiently specific terms so that, barring an influx of additional resources, the limited funds currently allocated for such services can be applied where most needed.

It is crucial to note that the great majority of children who died of homicide during the study period had no prior contact with child protective services. Most current systems of child protection target services to the victims of child abuse with the goal of minimizing the impact or negative consequences of abuse. At least two assumptions limit the applicability of this primarily tertiary prevention approach: first, that all abused children are detected, and second, that abuse is repetitive. Protection services will be successful only to the extent that

abused children are identified and given helpful services. However, we know that many maltreated children are not detected. Moreover, if some children are victims of single violent (and fatal) episodes, traditional social service intervention is not likely to have a preventive impact simply because such cases are not identified until it is too late.

In addition to exploring new mechanisms to detect children at risk of homicide, existing service delivery systems may serve a unique preventive function. One innovative program uses animal control and state humane officers, who have 24-hour staff and are often allowed to make unannounced visits to a home in which an animal was mistreated, to identify potentially abused children (K. White, Humane Society of the United States, personal communication, June 1993). However, the ability of this program to detect abused children and prevent further maltreatment (and perhaps homicide) remains to be tested.

These recommendations may be of assistance in addressing the homicide risk of some children and do not preclude other public health efforts to reduce homicides. Social, product, legislative, law enforcement, mental health, and other approaches are needed to meet the complex public health challenge presented by the homicide of children. □

Acknowledgments

This work was supported by the Southern California Injury Prevention Research Center through the Centers for Disease Control grant R49/CCR903622.

Preliminary findings were presented at the American Public Health Association Convention, November 10–14, 1991, in Atlanta, Ga.

We would like to thank former Chief Darryl Gates and Capt William Gartland of the Los Angeles Police Department for making the homicide case summaries available. We also thank the Los Angeles County Department of Children's Services for allowing access to records, and we extend special appreciation to the many people there who helped make data collection a reality during a time of transition. Thank you to Superior Court Judge Paul Boland who, in his former role as a juvenile court judge, made this research possible; and to Barbara Ahmad, Hal Morgenstern, Dawn Upchurch, and Daniel Webster for their helpful comments. Special thanks to Haikang Shen for data analytic work.

References

1. National Center for Health Statistics. Advance report of final mortality statistics, 1990. *Month Vital Stat Rep.* January 7, 1993;41:21 (ranking determined from data provided in Table 7).
2. Christoffel KK. Homicide in childhood: a public health problem in need of attention. *Am J Public Health.* 1984;74:68–70.
3. US Bureau of the Census. Characteristics of American children and youth: 1980. *Current Population Reports.* 1982;114:23.
4. Kellermann AL, Reay DT. Protection or peril? An analysis of firearm-related deaths in the home. *N Engl J Med.* 1988;314:1557–1560.
5. Sloan JH, Kellermann AL, Reay DT, et al. Handgun regulations, crime, assaults, and homicide. A tale of two cities. *N Engl J Med.* 1988;319:1256–1262.
6. Wintemute GJ, Teret SP, Kraus JF, Wright MW. The choice of weapons in firearm suicides. *Am J Public Health.* 1988;78:824–826.
7. Abel EL. Childhood homicide in Erie County, New York. *Pediatrics.* 1986;77: 709–713.
8. Blaser MJ. Epidemiologic characteristics of child homicides in Atlanta, 1970–1980. *Pediatrician.* 1983–1985;12:63–67.
9. Crittenden PM, Craig SE. Developmental trends in the nature of child homicide. *J Interpersonal Violence.* 1990;5:202–216.
10. Fingerhut LA, Kleinman JC, Godfrey E, Rosenberg H. Firearm mortality among children, youth, and young adults 1–34 years of age, trends and current status: United States, 1979–1988. *Month Vital Stat Rep.* 1991;39(11, suppl).
11. Fingerhut LA. Firearm mortality among children, youth, and young adults 1–34 years of age, trends and current status: United States, 1985–1990. *Adv Data Vital Health Stat.* 1993; no. 231.
12. Goetting A. Child victims of homicide: a portrait of their killers and the circumstances of their deaths. *Violence and Victims.* 1990;5:287–293.
13. Jason J. Child homicide spectrum. *Am J Dis Child.* 1983;137:578–581.
14. Muscat JE. Characteristics of childhood homicide in Ohio, 1974–1984. *Am J Public Health.* 1988;78:822–824.
15. Sorenson SB, Peterson JG, Richardson BA. Child homicide in Los Angeles: an examination of a decade of deaths. In: Gefner R, Sorenson SB, Lundberg-Love P, eds. *Violence at Home: Current Issues in Spousal Battering and Child Maltreatment.* Binghamton, NY: Haworth Press; forthcoming.
16. Copeland AR. Homicide in childhood. The Metro-Dade County experience from 1956 to 1982. *Am J Med Pathol.* 1985;6:21–24.
17. Gelles RJ. Physical violence, child abuse, and child homicide: a continuum or distinct behaviors? *Hum Nature.* 1991;2: 59–72.
18. Hollander N. Physical abuse as a predictor of child homicide. *Tex Med.* 1986;82: 21–23.
19. Breslow NE. Odds ratio estimators when the data are sparse. *Biometrika.* 1981;68: 73–84.
20. Breslow NE, Day NE. *Statistical Methods in Cancer Research.* Vol 1: *The Analysis of Case-Control Studies.* Lyon, France: International Agency for Research on Cancer; 1980.
21. Bass B, Wyatt GE, Powell GJ. *The Afro-American Family.* New York, NY: Grune & Stratton; 1982.
22. Sabotta EE, Davis RL. Fatality after report to a child abuse registry in Washington state, 1973–1986. *Child Abuse Negl.* 1992;16:627–635.
23. Durfee MJ, Gellert GA, Tilton-Durfee D. Origins and clinical relevance of child death review teams. *JAMA.* 1992;267: 3172–3175.
24. Ewigman B, Kivlahan C, Land G. The Missouri child fatality study: underreporting of maltreatment fatalities among children younger than five years of age, 1983 through 1986. *Pediatrics.* 1993;91: 330–337.
25. Schloesser P, Pierpont J, Poertner J. Active surveillance of child abuse fatalities. *Child Abuse Negl.* 1992;16:3–10.
26. Haddon W Jr, Ellison AE, Carroll RE. Skiing injuries. Epidemiologic study. *Public Health Rep.* 1962;11:975–991.
27. Christoffel KK, Liu K, Stamler J. Epidemiology of fatal child abuse: international mortality data. *J Chronic Dis.* 1981;34:57–64.
28. Gartner R. The victims of homicide. A temporal and cross-national comparison. *Am Sociol Rev.* 1990;55:92–106.
29. Kempe CH. Approaches to preventing child abuse. The health visitors concept. *Am J Dis Child.* 1976;130:941–947.
30. Olds DL, Henderson CR, Chamberlin R, Tatelbaum R. Preventing child abuse and neglect: a randomized trial of nurse home visitation. *Pediatrics.* 1986;78:65–78.
31. Frankel H. Family-centered, home-based services in child protection: a review of the research. *Soc Serv Rev.* 1988;61:137–157.
32. Winpisinger KA, Hopkins RS, Indian RW, Hostetler JR. Risk factors for childhood homicides in Ohio: a birth certificate–based case-control study. *Am J Public Health.* 1991;81:1052–1054.

Factors That Influence Clinicians' Assessment and Management of Family Violence

Virginia P. Tilden, DNSc, RN, Terri A. Schmidt, MD, Barbara J. Limandri, DNSc, RN, Gary T. Chiodo, DMD, Michael J. Garland, DScRel, and Peggy A. Loveless, MSW, LCSW

ABSTRACT

Objectives. High rates of family violence and low rates of detection, report, and therapeutic intervention by health professionals are well documented. This study was undertaken to determine what factors influence clinicians' decision making about identifying abuse and intervening with victims.

Methods. Survey data about clinicians' experiences with and attitudes toward family violence were gathered by mailed questionnaire from a random sample of practicing clinicians in six disciplines (n = 1521).

Results. Data showed similarities within and wide differences among three groups of subjects: dentists/dental hygienists, nurses/physicians, and psychologists/social workers. Overall, a third of subjects reported having received no educational content on child, spouse, or elder abuse in their professional training programs. Subjects with education on the topic more commonly suspected abuse in their patients than those without; among all subjects, spouse abuse was suspected more often than child abuse while elder abuse was suspected infrequently. Significant numbers of subjects did not view themselves as responsible for dealing with problems of family violence. Subjects indicated low confidence in and low compliance with mandatory reporting laws.

Conclusions. There is a need for educators to expand curricula on family violence and for legislators to reexamine mandatory reporting laws. (*Am J Public Health.* 1994; 84:628–633)

Introduction

Recent reports in the medical literature document that the incidence of family violence is much greater than has been assumed in the past and that health professionals tend not to screen patients for abuse.[1–4] Statistics indicate that between one fourth and one half of all intimate couples are violent at some time,[5] and that, annually within families, 3 to 4 million women are battered,[4] 1.5 million children are seriously abused, another 5 million children are maltreated,[6,7] and 2.5 million elders are abused or neglected.[8] These documented rates are likely to be underestimates because surveys typically do not include non-English-speaking persons, the very poor, those without telephones, and those in hospitals or other institutions.[4]

Victims of abuse present for treatment in clinical settings—both for their injuries and for health problems related to stress (e.g., anxiety, insomnia)—in numbers much greater than health professionals believe. For example, one third of women who visit hospital emergency departments do so because of symptoms related to ongoing abuse from their intimate partner,[9–11] and one sixth of pregnant women in public prenatal clinics have been abused within the past year.[12] Characteristically, victims of family violence do not disclose their abuse.

Thus, although health professionals often treat abused patients, in the vast majority of treatment situations they fail to suspect abuse.[9,10,13–15] When they do suspect it, the degree to which their responses are therapeutic or even compliant with state mandatory reporting laws varies widely.[16] The failure of clinicians to report suspected cases is well documented[16–26] and has been attributed to numerous causes, such as uncertainty about the clinical evidence, misgivings about the actions of protective services, and fear of involvement in court proceedings.[19,23,27–33]

While there has been some discipline-related empirical research on factors related to clinicians' unwillingness to detect and intervene in abuse,[1] there are little comparative data among the health professions on such factors. The aim of this study, therefore, was to explore factors, such as gender and training, that are likely to influence clinicians' assessment and management of abuse victims, and to examine factors within and among the disciplines of dentistry, dental hygiene, medicine, nursing, psychology, and social work. Practitioners within these six clinical disciplines have the opportunity and the responsibility for making clinical judgments about abused patients. Reported here are results of a mailed survey that was part of a larger study that included interviews with practitioners to explore their ethical reasoning when faced with evidence of domestic violence.

Methods

A mail survey was designed by the investigators to measure factors related

The authors are with Oregon Health Sciences University, Portland, Ore.

Requests for reprints should be sent to Virginia P. Tilden, DNSc, RN, Office of Research SN-ORD, Oregon Health Sciences University, 3181 SW Sam Jackson Park Rd, Portland, OR 97201.

This paper was accepted September 20, 1993.

TABLE 1—Subjects by Gender, Residence, and Year of Graduation from Professional Training

	Dental Hygienists (n = 275)	Dentists (n = 250)	Nurses (n = 241)	Physicians (n = 221)	Psychologists (n = 261)	Social Workers (n = 273)	Total (n = 1521)
Gender, %							
Male	1	92	8	75	61	32	44
Female	99	8	92	25	39	68	56
Residence, %							
Rural	26	26	31	26	15	20	24
Urban	74	74	69	74	85	80	76
Year of graduation							
Median	1979	1973	1978	1978	1978	1977	1977
Range	1943–1989	1942–1990	1946–1990	1946–1990	1951–1989	1942–1987	1942–1990

TABLE 2—Percentage of Subjects Who Recalled Any Educational Content in Their Professional Training Programs on Three Types of Domestic Violence

	Dental Hygienists (n = 274)	Dentists (n = 249)	Nurses (n = 240)	Physicians (n = 221)	Psychologists (n = 261)	Social Workers (n = 271)	Total (n = 1516)
Child abuse	45	45	75	69	73	84	64
Spouse abuse	14	15	44	40	59	67	40
Elder abuse	17	11	43	24	25	45	27

to clinicians' assessment and management of suspected abuse. Literature related to decision making about abuse[21,24–30] guided the development of survey items. The four-page survey included (1) demographic items (respondent's discipline, gender, and year of graduation); (2) items that tapped subjects' recall of the amount of educational content about family violence they received in their professional training programs (scored as none, a little, a moderate amount, or a great deal); (3) frequency with which subjects suspect that their patients have been abused (scored as never, occasionally, a moderate amount, or frequently); (4) subjects' typical responses or interventions when abuse is suspected (from a checklist of typical interventions); and (5) attitude items regarding both their extent of responsibility in dealing with domestic violence and mandated reporting. The survey was pilot tested initially with two or three clinicians in each of the six disciplines and was then revised. The revised form was retested with a sample of 30 clinicians, 5 in each discipline, selected from a health sciences university directory. Twenty-seven responses were received, and inspection of them led to final changes in the survey. Time to complete the survey averaged 10 minutes in pilot tests.

Following University Institutional Review Board approval, the survey was mailed to 2100 health care professionals (350 practitioners in each of six health care disciplines) who represented a random sample of professionals with licenses to practice in a northwestern state. Surveys were mailed under a cover letter that was tailored slightly to each discipline and signed by the member of the research team from that discipline. Returns were tracked by a research assistant. Nonresponders at 2 weeks were sent a reminder postcard; 4 weeks after the initial mailing, nonresponders were sent a second survey. Mailing lists were then destroyed.

Data were analyzed with the Crunch 4.0 statistical software package. Descriptive (central tendencies, frequencies) and inferential (correlation, chi-square, logistic regression) statistics were computed.

Results

Of the 2100 questionnaires mailed, 1588 were completed and usable, 53 were returned as undeliverable, and 3 were returned as addressee deceased. The overall response rate was 78%; however, response rates for the specific disciplines varied (dental hygienists, 83%; social workers, 82%; psychologists, 81%; nurses, 78%; dentists, 78%; physicians, 69%). Sixty-seven people reported retirement and their responses were excluded, leaving 1521 questionnaires for analysis. Surveys contained little missing data, probably owing to the extensive pilot testing that went into formatting items and response choices to minimize the burden of responding. The small amount of item-level missing data that occurred is reflected in the slightly different sample numbers reported on the individual items as noted in the tables.

Sample Characteristics

Sample characteristics are presented in Table 1. The overall sample included a slightly higher proportion of women than men although, as expected, the gender proportions in individual disciplines were quite skewed. The median year of graduation from their respective professional training was 1977, and time since graduation ranged from 1 to 49 years, with a median of 14 years. Overall, 76% of survey subjects came from urban counties, which reflected regional population demographics in

TABLE 3—Percentage of Subjects Who Reported Commonly Suspecting Physical or Sexual Abuse among Patients

	Dental Hygienists (n = 271)	Dentists (n = 247)	Nurses (n = 236)	Physicians (n = 218)	Psychologists (n = 260)	Social Workers (n = 269)
Child abuse						
Physical	0.4	2.0	9.3	8.8	24.8	40.0
Sexual	0.4	1.2	7.7	8.9	30.9	43.2
Spouse abuse						
Physical	3.7	6.0	18.2	22.7	46.2	53.3
Sexual	0.8	2.9	7.9	9.2	23.9	31.2
Elder abuse						
Physical	0.7	0.0	10.7	3.2	2.3	8.5
Sexual	0.0	0.0	2.1	0.5	3.9	3.8

TABLE 4—Adjusted Odds Ratios (ORs) and 95% Confidence Intervals (CIs) for the Effect of Educational Content on Suspicion of Abuse, with Gender, Graduation Year, and Discipline Controlled

Form of Suspected Abuse	Adjusted OR	95% CI
Child		
Physical (n = 1477)	1.68	1.09, 2.58
Sexual (n = 1464)	1.66	1.09, 2.54
Spouse		
Physical (n = 1470)	1.33	1.00, 1.77
Sexual (n = 1445)	1.39	0.98, 1.98
Elder		
Physical (n = 1453)	2.10	1.20, 3.69
Sexual (n = 1441)	1.24	0.53, 2.93

that 75% of the population in the northwest region resides in urban counties. In subsequent analyses, however, there were no statistically significant differences between urban and rural subjects on any of the study variables.

Education

The extent of subjects' formal education (i.e., excluding postgraduate continuing education) on the topic of domestic violence varied widely according to discipline and year of graduation. Table 2 shows the percentage of subjects reporting any (i.e., a little, a moderate amount, or a great deal) educational content on child, spouse, and elder abuse in their professional training programs.

Combining groups, 64% of subjects reported education in child abuse, 40% reported education in spouse abuse, and 27% reported education in elder abuse. Of the total sample, 33% (n = 499) reported no education in any of these three types of abuse. Examining subjects by discipline, social work had the highest proportion of subjects reporting any education in each type of abuse, and dentists and dental hygienists had the smallest proportion. For all types of abuse, the percentage of subjects who reported any education increased as year of graduation became more recent.

Suspecting Abuse

Subjects were asked how often they suspected that a child, adult, or elderly patient had been a victim of physical or sexual abuse. Table 3 shows the percentage of subjects who indicated that they suspected abuse either a moderate amount or frequently, combined here as commonly. In general, the numbers of subjects who commonly suspected abuse were low in all disciplines except social work and psychology. Generally, subjects were most likely to suspect physical abuse of spouses, followed by physical or sexual abuse of children. Elder abuse, either physical or sexual, was not frequently suspected by any discipline.

Subjects who reported receiving any educational content on family violence in their training programs were more likely to indicate that they commonly suspected child, adult, and elder abuse, both physical and sexual, among their patients. For example, among those with education, 19.0% and 20.6% commonly suspected physical and sexual child abuse, respectively, while among those without education, the proportions were 6.6% ($P < .001$) and 7.2% ($P < .001$), respectively. This statistically significant effect of education on suspicion of abuse held true for spouse and elder abuse as well.

We used logistic regression to test the association between educational content on domestic violence (dichotomized as none vs any) and suspicion of abuse while controlling for the possible confounding variables of gender, year of graduation, and discipline. Results are shown in Table 4.

Education predicted more than one-and-a-half times the rate at which physical and sexual child abuse was suspected and more than twice the rate at which physical abuse of elders was suspected. However, education did not strongly influence suspicion of spouse abuse or sexual abuse of elders. It may be that sexual abuse of elders is relatively infrequent, which would thus limit the extent to which factors such as education would influence any suspicion of its occurring. Yet it is clear that education, entirely apart from other factors, positively affects the rate at which clinicians suspect child and physical elder abuse.

Interventions

We asked the clinicians the extent to which they view themselves as responsible for intervening in suspected domestic violence. Responses to the attitude item "Professionals in my discipline have as much responsibility to deal with problems of family violence as they do to deal with other clinical problems" indicated that, although the majority of clinicians agreed with the statement, remarkably large numbers in some disciplines, especially dental hygiene and dentistry, did not agree (dental hygienists, 45.9%; dentists, 47.3%; nurses, 13.3%; physicians, 15.4%; psychologists, 1.5%; social workers, 2.6%).

From a list of interventions, subjects were asked to choose their typical responses when they encounter an adult, child, or elder whom they suspect of having been abused. Table 5 presents the interventions selected most often by subjects in each discipline. Reporting was rarely the most common intervention for clinicians in any discipline even

TABLE 5—Most Frequently Selected Intervention[a] for Child, Spouse, and Elder Abuse

	Dental Hygienists (n = 275)	Dentists (n = 250)	Nurses (n = 241)	Physicians (n = 220)	Psychologists (n = 261)	Social Workers (n = 273)
Child abuse	Consult (39%)	Chart and observe (45%)	Consult (34%)	Report (39%)	Discuss with family (44%)	Report (59%)
Spouse abuse	Discuss with patient (26%)	Discuss with patient (39%)	Consult (51%)	Discuss with patient (75%)	Discuss with patient (77%)	Discuss with patient (78%)
Elder abuse	Chart and observe (21%)	Chart and observe (24%)	Consult (52%)	Discuss with patient (41%)	Discuss with patient (15%)	Discuss with patient (23%)

[a]Choices included (1) make a note in the chart and observe over time; (2) discuss the observations with the patient; (3) discuss the observations with family members; (4) consult with another professional; (5) refer the patient to social service/mental health service; (6) report/encourage patient to report to appropriate protective service agency or to police; (7) facilitate arrangements for safety; (8) elect not to intervene; or (9) other.

though mandated reporting laws for child and elder abuse have existed in the state since 1971 and 1981, respectively. Only among social workers and physicians was reporting selected most often and then only for child abuse.

In addition to these significant differences among disciplines in choice of intervention, gender affected the frequency with which subjects consulted with another professional. For all types of abuse (child, spouse, and elder), women were more likely than men to consult with another professional ($P < .001$ for each). Using logistic regression to test the effect of gender on choice of intervention while controlling for discipline, we found significant gender differences in some of the interventions selected for child abuse and spouse abuse but not for elder abuse. For child abuse, women were almost twice as likely as men to consult with another professional (adjusted odds ratio = 1.94; 95% confidence interval [CI] = 1.33, 2.82). Also, women were one and a half times more likely than men to report abuse to child protective services (adjusted odds ratio = 1.66; 95% CI = 1.05, 2.61). For spouse abuse, women were one-and-a-half times more likely than men to consult with another professional (adjusted odds ratio = 1.52; 95% CI = 1.08, 2.13), while men were one-and-a-half times more likely than women to discuss the observations with a family member (adjusted odds ratio = 1.61; 95% CI = 1.08, 2.44).

Next, we used logistic regression to examine the effect of educational content on choice of interventions while controlling for the effects of gender, year of graduation, and discipline. Education significantly predicted the choice of one intervention in child abuse (make a note in the chart and observe overtime [adjusted odds ratio = 1.42; 95% CI = 1.00, 2.03]); of three interventions in spouse abuse (make a note in the chart and observe [adjusted odds ratio = 1.41; 95% CI = 1.05, 1.89], discuss observations with family [adjusted odds ratio = 1.75; 95% CI = 1.19, 2.59], and consult with another professional [adjusted odds ratio = 1.58; 95% CI = 1.16, 2.15]); and of four interventions in elder abuse (make a note in the chart and observe [adjusted odds ratio = 2.4; 95% CI = 1.58, 3.63], discuss observations with patient [adjusted odds ratio = 1.51; 95% CI = 0.99, 2.28], discuss observations with family [adjusted odds ratio = 1.61; 95% CI = 1.06, 2.45], and report to appropriate protective service agency [adjusted odds ratio = 1.57; 95% CI = 1.00, 2.46]). An interesting feature of these findings is that formal educational content in family violence significantly predicted the reporting only of elder abuse.

TABLE 6—Subjects' Attitudes (%) toward Mandatory Reporting Laws for Child, Spouse, and Elder Abuse

	Dental Hygienists (n = 275)	Dentists (n = 251)	Nurses (n = 241)	Physicians (n = 221)	Psychologists (n = 261)	Social Workers (n = 273)
Child abuse: laws are effective	36	37	39	46	50	52
Spouse abuse: there should be a law	66	55	76	56	48	68
Elder abuse: laws are effective	33	32	35	35	39	38

Mandated Reporting: Compliance and Attitudes

Of those clinicians who indicated that their practices included children, the percentage who included reporting in the list of interventions they typically used varied (dental hygienists, 8.8%; dentists 20.9%; nurses, 44.4%; physicians, 74.5%; psychologists, 73%; social workers, 87.5%). Of those clinicians who indicated that their practices included the elderly, the percentage who included reporting among interventions they used also varied (dental hygienists, 6.5%; dentists, 12.2%; nurses, 32%; physicians, 45.5%; psychologists, 60.8%; social workers, 76.4%).

These relatively low rates of reporting are reflected in subjects' attitudes about the effectiveness of mandatory reporting, which are presented in Table 6. Only about one third of subjects in each discipline thought that mandatory reporting of suspected elder abuse is effective in dealing with the problem.

Approximately half of the physicians, psychologists, and social workers thought that mandatory reporting of suspected child abuse is effective in dealing with that problem, as did smaller percentages (37% to 39%) of dental hygienists, dentists, and nurses. Despite doubts about the effectiveness of mandatory reporting of suspected child and elder abuse, a majority of subjects in all disciplines except psychology thought there should also be a mandatory reporting law for spouse abuse; more women than men held this opinion as well (69% vs 53%, $P < .001$). We used logistic regression to assess the effect of educational content on subjects' attitudes about mandatory reporting laws while controlling for gender, year of graduation, and discipline. In this multivariate model, there were no significant effects of the independent variables on attitudes about reporting.

Discussion

Although it is difficult to isolate factors that relate to clinicians' responses to victims of family violence, certain trends in the data may be of value to educators and legislators. First, there are similarities within and wide differences among three groups of subjects: dentists/dental hygienists, nurses/physicians, and psychologists/social workers. This fact, when combined with the low rate of education on family violence in professional training programs, indicates that clinicians' responses may stem more from their disciplinary background than from knowledge about family violence. Given that clinicians in all six disciplines are about equally exposed to the public and equally responsible for detecting, reporting, and responding to violence, the lack of similarity in responses across disciplines and the discipline-specific patterns are disturbing.

A second disturbing finding is the high numbers of currently practicing clinicians without education about this highly prevalent health problem. Fully one third are without education in *any* of the three main types of domestic violence. Beyond that, approximately one third are without education in child abuse, two thirds are without education in spouse abuse, and three quarters are without education in elder abuse. Although the trend toward more education among more recent graduates is encouraging and consistent with the literature,[14] the findings nonetheless point to an enormous lack in professional training programs. That approximately one half of graduates since 1980 reported no education in spouse or elder abuse and a fifth reported no education in child abuse are startling findings, given that it was in the early 1980s that Surgeon General C. Everett Koop called for increased national attention to domestic violence as a major public health problem.[34]

The issue of education is especially important given the influence it exerts on the rates with which clinicians suspect abuse, an influence that is separate from that exerted by other variables, specifically gender, recency of graduation, and discipline. Although, the findings overall show low rates of suspecting abuse, clinicians with education do suspect abuse more commonly.

Authors in the recent literature have called for major revisions in child abuse curricula for medical[35–37] and dental students.[21,38] We found greater gaps in education about spouse and elder abuse, and therefore we expand that call to include even more extensive revisions in curricula for all health professionals focused on abuse in these areas. There appears to be a disproportionate emphasis in education on child abuse, probably because it presents a clearer model of victim and abuser. The result is that health professionals often lack an understanding of the more complicated dynamics associated with the abuse of adults, such as reasons why victims often stay in abusive situations. This may be a factor in clinicians' less therapeutic response to battered women, as noted in prior literature.[16]

A third disturbing finding is that large numbers of health professionals in each discipline believed abuse to be infrequent among their patients. These findings support those of prior studies.[9,10,13–15] Abuse is a hidden health care problem, and the unsuspecting stance of health professionals allows the problem to remain undetected much of the time.

Dentists and dental hygienists reported the least education in abuse, the least frequent rate of suspecting abuse, and the greatest proportion who do not see themselves as responsible for intervening into suspected abuse. We encourage dentistry and dental hygiene to recognize that injuries to the teeth, jaw, neck, and oral soft tissues are common indicators of abuse, and that dentists and hygienists often have the first opportunity to observe these symptoms.[39]

The question of how clinicians choose to intervene showed trends reflective of gender-based reasoning styles noted by others.[40] Gender influences how clinicians prefer to handle suspected child and spouse abuse cases, with women preferring the more collaborative approaches of consulting with others and reporting to others, and men preferring to handle things alone by discussing the problem with the family. Although we hesitate to place a value judgment on styles, we do call for increased collaboration and communication between health care team members in order to combat the isolation that tends to suppress suspicion of and intervention in cases of abuse.

Significant numbers of professionals do not view themselves as responsible for intervening, and those who do intervene tend not to report abuse to the appropriate authorities. This issue of attitudes about and compliance with mandatory reporting laws deserves a great deal of reflection. The vast majority of dental hygienists and dentists, over a half of the nurses, and a quarter of the physicians said they would *not* report a case of suspected child abuse to authorities; even greater numbers said this about suspected elder abuse. Although those with more education more commonly *suspect* abuse, education does not influence their *reporting* of abuse. Subjects indicated low confidence in mandatory reporting, which their noncompliance reflects. Despite this, a majority indicated a wish for a mandatory reporting law for suspected spouse abuse, reflecting the human tendency to wish for external solutions even in the face of negative evidence.

In our discussions with clinicians, many said the primary reason they avoid reporting is because doing so sets into motion a series of events that unfold out of their control and are often counter-therapeutic. Thus, clinicians may experience an ethical dilemma between their perceived duty to report and their low confidence in the system's ability to respond effectively. This untenable position likely contributes to the large numbers of health professionals who say they do not suspect abuse, given that denial is one solution to an ethical dilemma. State officials and legislators should consider the problem of underfunding and overloading of protective services vis-à-vis clinicians' low level of

confidence and their resulting noncompliance with statutes. Many clinicians would prefer that reporting be mandatory for nonclinical professionals (e.g., teachers) or nontreating clinicians (e.g., office nurses) but that treating clinicians be allowed a margin of discretion regarding the most therapeutic intervention.

Findings of this study are limited by the restriction of the sample to the northwestern region, to problems associated with retrospective recall inherent in surveys, and to the effects of history during data collection. Regarding the latter, subjects' responses may have been affected by problems within child protective services that were covered widely in the news media during the period of data collection. Even so, trends in the data are sufficiently strong to inform educators of the need for expanded curricula on family violence and to inform legislators of the need to reexamine mandatory reporting laws relative to available protective services. □

Acknowledgments

This study was funded by the Medical Research Foundation of Oregon.

The investigators are grateful to Michelle Felix, MN, RN, for assistance with data collection; to the Oregon Health Sciences University Center for Ethics in Health Care, and especially to Susan Tolle, MD, for support and manuscript review; and to the Oregon Health Sciences University Office of Research Development and Utilization, and especially to Jonathan Fields, MA, for assistance with data management and analysis.

References

1. Sugg NK, Inui T. Primary care physicians' response to domestic violence. *JAMA.* 1992;267:3157–3160.
2. Novello AC, Shosky J, Froehlke R. From the Surgeon General, US Public Health Service: a medical response to violence. *JAMA.* 1992;267:3007.
3. American Medical Association Council on Ethical and Judicial Affairs. Physicians and domestic violence: ethical considerations. *JAMA.* 1992;267:3190–3193.
4. American Medical Association Council on Scientific Affairs. Violence against women: relevance for medical practitioners. *JAMA.* 1992;267:3184–3189.
5. Straus MA, Gelles RJ. Societal change and change in family violence from 1975 to 1985 as revealed by two national surveys. *J Marriage Fam.* 1986;48:465–480.
6. Straus MA, Gelles RJ. How violent are American families? Estimates from the national family violence resurvey and other studies. In: Hotaling G, Finkelhor D, Kirkpatrick J, Straus M, eds. *Family Abuse and Its Consequences: New Directions for Research.* Newbury Park, Calif: Sage; 1988.
7. National Center on Child Abuse and Neglect. *Study Findings: Study of National Incidence and Prevalence of Child Abuse and Neglect.* Washington, DC: US Dept of Health and Human Services; 1988.
8. American Medical Association Council on Scientific Affairs. Elder abuse and neglect. *JAMA.* 1987;257:966–971.
9. Appleton W. The battered woman syndrome. *Ann Emerg Med.* 1980;9:84–91.
10. Tilden VP, Shepherd P. Increasing the rate of identification of battered women in an emergency department: use of a nursing protocol. *Res Nurs Health.* 1987; 10:209–215.
11. McLeer SV, Anwar R. A study of battered women presenting in an emergency department. *Am J Public Health.* 1989;79:65–66.
12. McFarlane J, Parker B, Koeken K, Bullock L. Assessing for abuse during pregnancy. *JAMA.* 1992;267:3176–3178.
13. Goldberg WG, Tomlanovich MC. Domestic violence victims in the emergency department. *JAMA.* 1984;251:3259–3264.
14. McLeer SV, Anwar RA, Herman S, Maquiling K. Education is not enough: a systems failure in protecting battered women. *Ann Emerg Med.* 1989;18:651–653.
15. Rounsaville BJ, Weissman MM. Battered women: a medical problem requiring detection. *Int J Psychiatry Med.* 1977;8:191–202.
16. Kurz D. Emergency department responses to battered women: resistance to medicalization. *Soc Probl.* 1987;34:69–81.
17. Brewer RA, Jones JS. Reporting elder abuse: limitations of statutes. *Ann Emerg Med.* 1989;18:1217–1221.
18. Clark-Daniels CL, Daniels RS, Baumhover LA. Abuse and neglect of the elderly: are emergency department personnel aware of mandatory reporting laws? *Ann Emerg Med.* 1990;19:970–977.
19. Finkel KC. The failure to report child abuse. *Am J Dis Child.* 1985;140:329–330.
20. Green FC. Child abuse and neglect. *Pediatr Clin North Am.* 1975;22:329–339.
21. Kassebaum DK, Dove SB, Cottone JA. Recognition and reporting of child abuse: a survey of dentists. *Gen Dent.* 1991;39: 159–162.
22. McDonald AE, Reece RM. Child abuse: problems of reporting. *Pediatr Clin North Am.* 1979;26:785–789.
23. Morris JL, Johnson CF, Clasen M. To report or not to report: physicians' attitudes toward discipline and child abuse. *Am J Dis Child.* 1985;139:194–197.
24. Phillips LR, Rempusheski VF. A decision-making model for diagnosing and intervening in elder abuse and neglect. *Nurs Res.* 1985;34:134–139.
25. Saulsbury FT, Campbell RE. Evaluation of child abuse reporting by physicians. *Am J Dis Child.* 1985;139:393–395.
26. Walker G, Jones JS, Krohmer J. To report or not to report: emergency services' response to elder abuse. *Prehosp Disaster Med.* 1991;6:380.
27. Couppe L. A retrospective analysis of the relationship between decision-making process and child abuse. *Child Abuse Negl.* 1983;7:363–366.
28. Daniels RS, Baumhover LA, Clark-Daniels CL. Physicians' mandatory reporting of elder abuse. *Gerontologist.* 1989;29: 321–327.
29. Kalichman SC, Craig ME, Follingstad DR. Mental health professionals and suspected cases of child abuse: an investigation of factors influencing reporting. *Community Ment Health J.* 1988;24:43–51.
30. Kalichman SC, Craig ME, Follingstad DR. Professionals' adherence to mandatory child abuse reporting laws: effects of responsibility attribution, confidence ratings, and situational factors. *Child Abuse Negl.* 1990;14:69–77.
31. Kim DS. How physicians respond to child maltreatment cases. *Health Soc Work.* 1986;11:95–106.
32. Silber TJ. Ethical issues in the treatment of the sexually abused adolescent. *Semin Adolesc Med.* 1987;3:29–31.
33. Winefield HR, Castell-McGregor SN. Experiences and views of general practitioners concerning sexually abused children. *Med J Aust.* 1986;145:311–313.
34. Koop CE. *Surgeon General's Workshop on Violence and Public Health.* Leesburg, Va: US Public Health Service, Office of Maternal and Child Health; 1985.
35. Alexander RC. Education of the physician in child abuse. *Pediatr Clin North Am.* 1990;37:971–987.
36. Dubowitz H. Pediatrician's role in preventing child maltreatment. *Pediatr Clin North Am.* 1990;37:989–1002.
37. Krugman RD. Future role of the pediatrician in child abuse and neglect. *Pediatr Clin North Am.* 1990;37:1003–1011.
38. Posnick WR, Donly KJ. Instruction in child abuse and neglect in the predoctoral curriculum. *J Dent Educ.* 1990;54: 158–159.
39. Chiodo GT, Rosenstein DI. Child abuse in Oregon: the dentist's responsibility. *J Oregon Dent Assoc.* 1984;54:21–24.
40. Gilligan C. *In a Different Voice.* Cambridge, Mass: Harvard University Press; 1982.

Suicidal Ideation and Attempts among Chemically Dependent Adolescents

Eva Y. Deykin, DrPH, and Stephen L. Buka, ScD

ABSTRACT

Objectives. Suicidal ideation and attempts were examined in a population of chemically dependent adolescents, a group at high risk of self-destructive behavior.

Methods. The prevalence and correlates of suicidality and of major depressive disorder were assessed by the Diagnostic Interview Schedule and a structured family and social history interview with 300 addicts aged 15 through 19 years.

Results. Suicidal ideation was reported by 31% to 75% of the subjects and suicide attempts were reported by 28% to 61%, with females predominating. Thoughts of suicide combined with prolonged thoughts of death in general and a desire to be dead were highly associated with suicide attempts. Exposure to physical or sexual abuse was associated with a significantly increased risk of suicide attempts for males but not for females.

Conclusions. The probability of a suicide attempt increases when thoughts of suicide coincide with morbid ideation of extended duration, suggesting that risk assessment should be based on duration as well as presence of morbid thoughts. Substance abuse treatment requires an assessment of suicidal potential and counseling for those whose potential is high, with special attention to males exposed to abuse. (*Am J Public Health.* 1994;84:634–639)

Introduction

The dramatic rise in suicide among adolescents over the past quarter-century has not been matched by similar changes in other age groups. During the period 1960 through 1988, when the rate of suicide for persons aged 15 through 24 more than doubled from 5.2 to 13.2 per 100 000, the rate of suicide among those aged 45 through 64 decreased from 20.7 to 14.6 per 100 000 and the rate among those aged 75 through 84 decreased from 27.9 to 25.9 per 100 000.[1] Furthermore, a population study by Murphy and Wetzel[2] indicated that in the past 40 years, each cohort entering adolescence had a higher rate of suicide than the preceding cohort and this difference increased as each cohort aged through adolescence. This disparity among age groups suggests that the factors responsible for the increase in adolescent suicide either preferentially affect the young or are absent at older ages. Among the influences thought to have particular relevance for suicidal behavior in adolescence are increased competition for diminished resources,[3] exposure to child abuse,[4,5] high incidence of family fragmentation,[6–8] elevated occurrence of depressive illness in cohorts born after 1935,[9] greater availability of handguns,[10–14] and increased use of alcohol and other drugs.[10–15]

Substance use may be especially important in the development of suicidal and other self-destructive behaviors. Mind-altering substances tend to disinhibit impulsivity and facilitate the expression of suicidal impulses that might otherwise have been contained. In addition, substance abuse can result in social isolation and increase the risk of depressive mood by distancing an adolescent from family or by diminishing his or her opportunities because of school failure, lost jobs, or legal involvement. Lastly, chemical dependence may place the adolescent in close contact with drug dealers and users who have easy access to handguns. The combination of substance abuse and availability of handguns poses an especially high-risk situation.

A substance-abusing population is particularly informative for the study of adolescent suicidal behavior and ideations. Self-destructive ideation, behavior, and suggested risk factors are common among this population,[15] making it a useful group for empirical inquiry. Furthermore, this is a population with identified service needs that would benefit from improved prediction and treatment methods. Lastly, focusing on subjects at the extreme end of the substance-abusing spectrum, results from this population could provide useful information for preventive strategies among adolescent substance abusers who have not yet reached the level of chemical dependence.

The purpose of this study was threefold: first, to ascertain the prevalence of various forms of morbid/suicidal ideation and of suicide attempts among a group of adolescents who meet DSM-III-R criteria for chemical dependence[16]; second, to examine the relationship of various types of morbid ideation to the occurrence of suicidal attempts; and third, to assess selected social and

The authors are with the Department of Maternal and Child Health, Harvard School of Public Health, Boston, Mass.

Requests for reprints should be sent to Eva Y. Deykin, DrPH, Department of Maternal and Child Health, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115.

This paper was accepted June 7, 1993.

demographic characteristics, including a history of victimization, as possible predictors of suicide attempts.

Sample and Settings

The study sample consisted of 300 adolescents (224 males, 76 females) aged 15 through 19 who were receiving residential treatment for their dependence on alcohol or other drugs. This sample was assembled during the course of a larger study of the prevalence and correlates of psychiatric and physical comorbidity among chemically dependent adolescents. Subjects were enrolled in the study from seven of eight Massachusetts treatment sites that were closely affiliated with the Massachusetts Department of Public Health and that had residential treatment programs designed for adolescents. The eighth treatment center chose not to participate in the study. The cost of care for most of the adolescents at these facilities was covered by Medicaid or by state funds through the Department of Public Health. The eligible study sample thus included consecutively admitted adolescents receiving state support for residential treatment for chemical dependence.

The seven facilities provided treatment modeled after the Alcoholics Anonymous 12-step program, which emphasizes clearly defined rules and expectations in conjunction with graded levels of responsibility and freedom. The treatment focused on behavior modification rather than on psychiatric morbidity, and, as a matter of policy, the treatment sites refused admission to youths who had florid psychiatric symptoms. We enrolled willing subjects into the study between the 1st and 4th weeks of treatment. For adolescents who were younger than 18 years and who were not medically emancipated, we obtained signed consent from their parents as well. The institutional review boards of the Harvard School of Public Health and the Massachusetts Department of Public Health and the directors of all seven treatment sites approved the study protocol, the methods of obtaining informed consent, and the content of the consent statement. Eligible subjects who refused to participate (18%) were almost identical to enrolled subjects with respect to age, gender, and race, suggesting that the study sample is representative of eligible adolescents at the treatment sites.

TABLE 1—Sociodemographic Characteristics of Study Subjects

	Males (n = 224)	Females (n = 76)	χ^{2a}	P
Mean age, y	16.8	16.2		
White, %	83	81	0.30	.58
Grade retention ≥ 2 y, %	32	20	4.24	.04
Socioeconomic status[b]				
Class I, %	4	9	4.45 (*df* = 4)	.35
Class II, %	37	44		
Class III, %	31	28		
Class IV, %	24	16		
Class V, %	4	4		
Family ever received welfare, %	59	69	2.50	.11
Met DSM-III-R criteria for chemical dependence				
Alcohol only, %	4	12	9.15 (*df* = 2)	.01
Drugs only, %	5	11		
Both alcohol and drugs, %	91	78		
Either parent treated for psychological or emotional problem, %	40	65	12.54	<.0001
Either parent treated for alcohol problem, %	68	75	0.30	.58
Either parent treated for drug problem, %	47	47	0.01	.94
Victimization				
Child abuse report filed, %	20	43	14.10	<.0001
Victim of physical abuse, %	41	55	4.07	<.04
Victim of sexual abuse, %	10	72	110.37	<.0001
Any of the above 3 items, %	41	77	31.74	<.0001
Removed from home by protective authorities, %	20	41	12.66	<.0001

[a] *df* = 1 unless otherwise noted.
[b] Determined by the Hollingshead Index.[21] Class I = most advantaged; Class V = least advantaged.

Methods

We obtained information on suicidal ideation and suicidal attempts by means of the Diagnostic Interview Schedule (DIS), Version III-R,[17] a validated, structured interview that ascertains the prevalence of psychiatric disorders according to DSM-III-R criteria. Interviewers were college graduates who had previously worked with adolescents and who successfully completed two weeks of training in the administration and coding of the DIS. Training included background on the structure and purpose of the DIS, mock interviewing and coding, homework exercises, and finally actual interviews with the trainer present. Regular meetings were held throughout the study period to review interview techniques and to ensure that interviewers did not stray from the established protocol.

The section of the DIS that assesses major depressive disorder includes questions on the occurrence of three forms of morbid or suicidal ideation, as well as on the occurrence of a suicide attempt. The questions were as follows: (1) "Has there ever been a period of *2 weeks or longer* when you thought a lot about death, your own, someone else's, or death in general?" (2) "Has there ever been a period of *2 weeks or longer* when you desired to be dead?" (3) "Have you ever thought of committing suicide?" (4) "Have you ever attempted suicide?" The first two questions differed from the last two in that they specified a duration, whereas the last two inquired about any occurrence. Each question elicited a yes or no answer. Data on sociodemographic characteristics and on victimization were obtained by a structured interview also administered at the time of the subject's enrollment into the study. Information on parents' substance abuse and psychological problems was obtained from the subjects and not by direct examination of the parents. Data on parents' psychological problems were elicited by the follow-

TABLE 2—Prevalence of Suicidal Thoughts and Suicide Attempts, by Gender

	Males (n = 224)		Females (n = 76)		Relative Risk[a]	95% Confidence Interval	P
	No.	%	No.	%			
Thoughts of death in general	105	46.8	53	69.7	1.8	1.2,2.5	.001
Desire to be dead	69	30.8	44	57.9	1.6	1.3,2.0	<.001
Thoughts of suicide	121	40.3	57	75.0	1.8	1.2,2.8	.001
Suicide attempt	63	28.1	46	60.5	1.8	1.4,2.4	<.001

[a]Relative risk calculated with males as the reference group.

TABLE 3—Suicide Attempts by Number of Thoughts Experienced, by Gender

	Males (n = 224)			Females (n = 76)				
Number of Thoughts	No. Experiencing Ideation	Suicide Attempts		No. Experiencing Ideation	Suicide Attempts		χ^{2a}	P
		No.	%		No.	%		
None	72	2	2.7	10	0	0.0	...	...
One	65	15	23.0	16	6	37.5	2.36	.15
Two	31	11	35.4	12	12	41.6	0.14	.72
Three	56	35	62.5	38	35	92.1	10.44	<.001

[a]df = 1 (comparing males vs females).

ing questions: "Has your father (mother) ever had a psychological or emotional problem? Has he/she ever received treatment for this problem?" Only parents who were reported as having received treatment were classified as having a psychological difficulty. Similarly, parental alcohol and drug use was determined by the following question: "Has your father (mother) ever been treated for his/her use of alcohol (drugs)?" We assessed victimization by subjects' answers to four questions: (1) whether a report of child abuse or neglect had ever been filed with a child protective agency on behalf of the subject or any sibling; (2) whether the subject had ever experienced physical abuse; (3) whether the subject had ever been subjected to sexual abuse; and (4) whether the subject had ever been removed from the home by authorities of the child protective system. Each question elicited a yes or no answer.

Data were analyzed by SAS.[18] Associations of suicide attempts with morbid ideation, gender, and history of victimization were assessed by point estimates and confidence intervals of relative risks. All analyses were done separately for each gender and then for the total sample, controlling for gender with the Mantel-Haenszel summary relative risk statistic.[19] *P* values were computed by the Cochran-Mantel-Haenszel statistic.[20]

Results

Characteristics of Subjects

Table 1 shows the personal characteristics of the study subjects for each gender. Despite representation of all socioeconomic groups as determined by the Hollingshead Four-Factor Index,[21] the sample was largely a disadvantaged one characterized by a high degree of welfare recipiency, parental alcohol and drug abuse, parental emotional or psychological problems, and a high prevalence of victimization among the subjects. Males were significantly more likely than females to be at least two grades below their age-appropriate level and to meet DSM-III-R criteria for dependence on both alcohol and drugs. Females were more likely to report at least one parent with psychological or emotional problems and to have been victimized through physical and sexual abuse.

Prevalence of Morbid Ideation and Suicide Attempts

Table 2 shows the reported prevalence of the three types of morbid ideation and of suicide attempts. Females were almost twice as likely as males to have experienced each type of ideation and were more than twice as likely to have made a suicide attempt. Table 3 shows the proportion of subjects who had made a suicide attempt among those reporting no, one, two, or three types of morbid ideation. For both genders, the proportion of subjects attempting suicide increased with the number of thoughts experienced. The increase was slightly but not significantly more marked among females. Only in the extreme group of those who had experienced three types of ideation were females significantly more likely to be suicide attempters. These data suggest that both gender and the number of types of ideation play a role in the development of a suicide attempt but that female gender has its greatest impact only when all three types of ideation are present.

Type of Morbid Ideation as a Predictor of a Suicide Attempt

The most sensitive predictor of a suicide attempt was having experienced a thought of suicide, which was present among 98% of the suicide attempters. However, since the thought of suicide was also reported by 38% of nonattempters, this type of ideation was not specific to attempters. For the entire sample, the risk of a suicide attempt among those who experienced a thought of suicide was only 59%. We therefore investigated whether the risk of a suicide attempt varied when a thought of suicide occurred with and without the other two types of morbid ideation. When a thought of suicide was experienced with neither of the other two types of ideation, the probability of an attempt was 44.4%. It remained at approximately this level when it occurred only with thoughts of death in general (40.0%) and when it was present with only a desire to be dead (42.9%). However, when a thought of suicide co-occurred with both of the other types of ideation, the risk of a suicide attempt was elevated to 74.5%. The data in Table 4 show that although a thought of suicide, either alone or in

combination with one other type of ideation, elevates the risk of suicide to between 16 and 18 times that of subjects without such a thought, the combination of all three types of ideation nearly doubles the elevated risk to 30-fold. These data indicate that although a thought of suicide is in itself an important risk factor for a suicide attempt, its impact is greatly magnified when it is joined by the other two types of ideation.

Since depressive illness is both a known cause of suicide and a concomitant of morbid thoughts, it was necessary to control for depressive illness to obtain an unconfounded measure of the association of morbid thoughts with suicide attempt. Lifetime prevalence of major depressive disorder was obtained by means of the DIS according to DSM-III-R criteria and was found to be positive in 78 of the 300 study subjects (28%). Table 5 shows that when major depressive disorder was controlled, the association of the three types of ideation with attempted suicide still held, indicating that these thoughts constituted an independent risk factor for attempted suicide apart from that imposed by depressive illness. In fact, the combination of the three types of ideation proved to be a better screen for a suicide attempt than the diagnosis of major depressive disorder. When the diagnosis of major depressive disorder was used as the screen for attempted suicide, its sensitivity was 46.7%, its specificity was 86%, and its predictive value was 65.4%. In comparison, the combination of the three types of ideation had values of 64.2%, 87.2%, and 74.5%, respectively, for the same screening parameters.

Suicidal Behavior and Victimization

Table 6 presents the association between attempted suicide and various aspects of victimization. The prevalence of all four indicators of victimization was very high in the total sample, ranging from a low of 25% for removal from the home to a high of 45% for physical abuse (see Table 1). All aspects of victimization were more common among females; the greatest gender difference was found for sexual abuse (72% of females vs 10% of males).

For females, exposure to victimization did not increase the risk of a suicide attempt. However, males who had been exposed to physical or sexual abuse or who stated that a report of abuse or neglect had been filed with authorities were between 1.8 and 2.4 times more likely than males without such exposures to have made a suicide attempt. In all instances, this elevation of risk for males was statistically significant. Finally, removal from the home by protective authorities was not associated with a suicide attempt in either gender or for the total sample. However, since removal from the home may occur for a variety of reasons, we reassessed its association with suicide attempt only among those 147 subjects who had experienced some form of victimization. Surprisingly, in this restricted sample, removal from the home was not significantly associated with attempted suicide. Those who had been removed from home had only a 33% greater risk of having made a suicide attempt ($\chi^2 = 2.72$, $P = .10$).

TABLE 4—Risk of Attempted Suicide, by Type of Morbid Ideation

	n	% Attempting Suicide	Relative Risk[a]	95% Confidence Interval	P
Thought of suicide only					
Yes	45	44.4	18.1	5.6, 57.9	<.001
No	122	2.5			
Thought of suicide plus thoughts of death in general					
Yes	25	40.0	16.2	4.8, 54.8	<.001
No	122	2.5			
Thought of suicide plus desire to be dead					
Yes	14	42.9	17.4	4.8, 62.1	<.001
No	122	2.5			
Thought of suicide plus both other ideations					
Yes	94	74.5	30.9	9.8, 93.1	<.001
No	122	2.5			

[a]Reference group is subjects who have not experienced thoughts of suicide.

TABLE 5—Suicide Attempts in Relation to Number of Morbid Thoughts, by Presence or Absence of DSM-III-R Major Depressive Disorder

	Number of Thoughts	No. Subjects	Suicide Attempters	
			No.	%
No major depressive disorder (n = 219)	0	78	2	2.6
	1	72	19	26.4
	2	30	11	36.7
	3	39	26	66.7
Major depressive disorder (n = 78)	0	1	0	0.0
	1	9	2	22.2
	2	13	5	38.5
	3	55	44	80.0

Note. Owing to lack of information, three subjects could not be evaluated for major depressive disorder.

Discussion

The prevalence of suicidal ideation and of suicide attempts in this population of substance-dependent youths is between five and seven times that found in other cross-sectional studies of community samples.[22-24] Because of differences in methods of data collection, wording of questions, source of study subjects, and subjects' characteristics, comparisons among groups are difficult. However, a previous cross-sectional study

TABLE 6—Association of Victimization and Attempted Suicide

	n	% Attempting Suicide	Relative Risk	95% Confidence Interval	P
Report of abuse/neglect filed for subject or siblings					
Males					
Yes	44	45.5	1.8	1.2,2.7	.007
No	173	24.9			
Females					
Yes	30	63.3	0.9	0.5,1.6	.62
No	40	57.5			
Total					
Yes	74	52.7	1.4	1.1,1.9	.02
No	213	30.9			
History of physical abuse					
Males					
Yes	80	41.3	1.9	1.2,2.9	.003
No	115	21.7			
Females					
Yes	36	63.9	1.2	0.8,1.9	.32
No	29	51.7			
Total					
Yes	116	33.7	1.6	1.1,2.2	.003
No	144	46.8			
History of sexual abuse					
Males					
Yes	23	60.9	2.4	1.6,3.7	<.001
No	198	24.8			
Females					
Yes	54	61.1	0.9	0.6,1.5	.94
No	21	61.9			
Total					
Yes	77	61.0	1.5	1.1,2.1	.009
No	219	28.3			
Removal from home by authorities					
Males					
Yes	43	30.2	1.1	0.6,1.8	.77
No	171	28.1			
Females					
Yes	30	56.7	0.9	0.6,1.3	.60
No	43	62.8			
Total					
Yes	73	41.1	1.0	0.7,1.4	.92
No	214	35.1			

Note. Totals are adjusted for gender.

of 424 college freshmen aged 16 through 19 years that used the same data collection instrument reported a prevalence of attempted suicide of 3.5%, or 1/10 the prevalence found in the present study.[23] In the college study, there were 40 students who met DSM-III criteria for substance abuse, including 4-(10%) who reported having made a suicide attempt. Among the 384 students who were not substance abusers, the suicide attempt rate was only 2.6%. The data from that study as well as those from the present study provide evidence for an increased risk of suicidal behavior among youth with severe substance involvement.

The cross-sectional nature of this study does not provide data on whether morbid ideation preceded or followed a suicide attempt. The strong associations between these variables indicate that they are not independent of one another, but the associations do not inform as to their relative occurrence. Conventional wisdom would suggest that morbid ideation is more likely to precede a suicide attempt rather than the other way around. However, the temporal sequence of these variables can be ascertained only through a longitudinal study of either a cohort or a case-referent design.

A clearer understanding of the interrelationship of suicidal thoughts with other types of morbid ideation has implications for both screening and clinical practice. Although a thought of suicide is strongly associated with a suicide attempt, the addition of thoughts of death, particularly a desire to be dead, potentiates the impact of a thought of suicide. This magnified risk may be due to the prolonged duration of the other two types of ideation. Self-destructive, morbid thoughts over an extended period of time provide clear indications for intervention activities. Because duration of ideation appears to enhance risk, we suggest that in the assessment of suicidal potential among substance-abusing adolescents one should elicit not just the presence of suicidal thoughts but their duration and tenacity.

A history of victimization significantly increased the risk of a suicide attempt only for males. The consistency of this finding across all indicators of victimization suggests that this may not be a chance result. Although it is possible that females can sustain victimization with less adverse emotional effects than males, we are not aware of any empirical findings suggesting such a trend. Rather, our finding could indicate that females were at very high risk for suicide attempts for reasons apart from victimization. In such a situation, victimization would not confer any observable added risk. Data supporting such an explanation are seen in Table 3, where for each level of suicidal ideation, the risk of attempted suicide is greater for females than for males. Furthermore, females were two and a half times more likely than males to have met criteria for major depressive disorder. Because of the excess of these risk factors among females it is possible that the deleterious impact of victimization is masked in this population of adolescent girls.

Lastly, the finding that removal from the home only slightly increases the probability of attempted suicide for maltreated subjects prompts three possible explanations. The first is that *severity* of victimization does not influ-

ence adverse emotional states associated with suicidal behavior. This explanation is not supported by the present data, and it is an unlikely explanation because studies have shown that children who have been severely abused have more deleterious outcomes than children who have experienced only mild abuse. A second explanation may be that removal from the home is not a good marker of severity of victimization. However, since the placement of a child outside the parental home is traumatic, is difficult to arrange, and involves the use of limited resources, protective authorities do not consider such action except in extreme cases of maltreatment. A third explanation for the slight increment in risk is that removal from the home is both a marker of severe victimization and, at the same time, potentially a restorative event. Children who have been removed from their homes may well have suffered more extreme abuse than those who were not removed, but they may also have had the advantage of a healing environment not available to children who experienced less severe victimization but who remained in their homes.

The results of this study indicate that youths who meet criteria for dependence on alcohol or drugs appear to be at exceptionally high risk of attempted suicide. Since a history of attempted suicide is known to be the most powerful predictor both of subsequent attempts and of completed suicide,[25,26] special vigilance is justified for substance-dependent adolescents, especially if a suicidal thought occurs in individuals who also have other morbid thoughts of long duration. The finding that morbid thoughts of extended duration combined with a thought of suicide magnify the risk of attempted suicide has not previously been quantified. Although this finding may apply only to adolescents at the most extreme end of substance abuse, further research is warranted among both adolescents who are substance abusers but not chemically dependent and adolescents who are not substance users. Similar findings in diverse populations of adolescents may suggest opportunities for more accurate community-based screening and suicide prevention efforts. □

Acknowledgments

This study was partly supported by a grant from the National Institute of Drug Abuse (RO1-DA-05834-02), by a grant from the William T. Grant Foundation (86-1079-86), and by support from the American Suicide Foundation.

The authors wish to thank Timothy Zeena, Kimberlie Kelley, and Michael Eisenstein for their help in the preparation of this manuscript.

References

1. *Vital Statistics of the United States, 1960–1988, Vol. 2; Mortality Part A.* Hyattsville, Md: National Center for Health Statistics; 1962 through 1990 vols.
2. Murphy GE, Wetzel RD. Suicide risk by birth cohort in the U.S. *Arch Gen Psychiatry.* 1980;37:519–523.
3. Holinger PC, Offer D. Prediction of adolescent suicide: a population model. *Am J Psychiatry.* 1982;139:302–306.
4. Deykin EY, Alpert JJ, McNamara J. A pilot study of the effect of exposure to child abuse or neglect on adolescent suicidal behavior. *Am J Psychiatry.* 1985; 142:1299–1303.
5. Deykin EY. The utility of emergency room data for record linkage in the study of adolescent suicidal behavior. *Suicide Life Threat Behav.* 1989;19:90–98.
6. Gispert M, Wheeler K, March L, Davis MS. Suicidal adolescents: factors and evaluation. *Adolescence.* 1985;20:753–762.
7. Adam KS, Lohrenz JS, Harper D, Streiner D. Early parental loss and suicidal ideation in university students. *Can J Psychiatry.* 1982;27:275–281.
8. Adam KS, Bouchoms A, Streiner D. Parental loss and family stability in attempted suicide. *Arch Gen Psychiatry.* 1982;39:1081–1085.
9. Klerman GL. The current age of youthful melancholia: evidence for increased depression among adolescents and young adults. *Br J Psychiatry.* 1988;152:4–14.
10. Brent D, Perper JA, Allman CJ. Alcohol, firearms and suicide among youth: temporal trends in Allegheny County, PA, 1960–1983. *JAMA.* 1987;257:3369–3372.
11. Markush RE, Bertolucci AA. Firearms and suicide in the U.S. *Am J Public Health.* 1984;74:123–127.
12. Loftin C, McDowall D, Wiersema B, Cottey T. Effects of restrictive licensing of handguns on homicide and suicide in the District of Columbia. *N Engl J Med.* 1991;325:1615–1620.
13. Boyd JH. The increasing rate of suicide by firearms. *N Engl J Med.* 1983;308:872–874.
14. Sloan JH, Rivara FP, Reagy DT, Ferris JAS, Path MRC, Kellerman AL. Firearm regulations and rates of suicide. *N Engl J Med.* 1990;322:369–373.
15. Felts WM, Chernier T, Barnes R. Drug use and suicide ideation and behavior among North Carolina public school students. *Am J Public Health.* 1992;82:870–872.
16. *DSM-III-R: Diagnostic and Statistical Manual for Mental Disorders.* 3rd ed., revised. Washington, DC: American Psychiatric Association; 1987.
17. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. *Arch Gen Psychiatry.* 1981;38:381–389.
18. *SAS/STAT User's Guide.* Release 6.03 edition. Cary, NC: SAS Institute Inc; 1988.
19. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22:718–748.
20. Kleinbaum DG, Cooper LL, Morgenstern H. *Epidemiologic Research.* Belmont, Calif: Lifetime Learning Publications; 1982:345.
21. Hollingshead AB. Four factor index of social position. New Haven, Conn: Yale University, Department of Sociology; 1975. Photocopied paper.
22. Vandivort DS, Locke BZ. Suicide ideation: its relation to depression, suicide and suicide attempt. *Suicide Life Threat Behav.* 1979;9:205–218.
23. Levy JC, Deykin EY. Suicidality, depression and substance abuse in adolescence. *Am J Psychiatry.* 1989;146:1462–1467.
24. Mishara BL, Baker H, Mishara TT. The frequency of suicide attempts: a retrospective approach to college students. *Am J Psychiatry.* 1976;133:841–844.
25. Shafi M, Carrigan S, Whittinghill JR, Derrick A. Psychological autopsy of completed suicide in children and adolescents. *Am J Psychiatry.* 1985;142:1061–1064.
26. Goldacre M, Hawton K. Repetition of self poisoning and subsequent death in adolescents who take overdoses. *Br J Psychiatry.* 1985;146:395–398.

Public Health Briefs

The Sexual Assault of Women at Work in Washington State, 1980 to 1989

Bruce H. Alexander, MS, Gary M. Franklin, MD, MPH, and Marsha E. Wolf, PhD

A B S T R A C T

Sexual assault in the workplace and the related factors have not been well studied. Workers' compensation claims data were used to describe work-related sexual assaults in Washington State between 1980 and 1989. Sixty-three cases of work-related rape were identified during this study period. The occupations of the rape victims were similar to occupations identified as high risk for other intentional injuries, and the rape incidents were characterized by isolation from the public and from coworkers. Estimates of industry-specific rates are presented. The identification, evaluation, and prevention of sexual assault and other workplace violence are discussed. (*Am J Public Health.* 1994;84:640–642)

Introduction

Work-related injuries from violence are recognized as an important workplace safety issue.[1,2] Approximately 12% of all occupational fatalities are homicides,[2,3] accounting for 40% of such fatalities among women and 10% among men.[4–6] Circumstances surrounding the occurrence of rape in the workplace and the characteristics of the victims are not well known, although some occupations may involve high risk.[7] Such information is important in order to develop and implement appropriate prevention strategies. We used workers' compensation claim data to describe rapes occurring in the workplace in Washington State from 1980 to 1989.

Methods

The Department of Labor and Industries is the sole regulator of workers' compensation coverage in Washington State. The state fund covers approximately two thirds of all nonfederal workers in Washington State; it excludes those working for self-insured establishments and workers covered by federal programs. The workers' compensation database contains diagnostic information (*International Classification of Diseases,* ninth revision,[8] codes and Z16.2 nature of injury codes[9]), medical procedure billing information, and limited information describing the injured worker, the type of establishment, and the occurrence of the injury. A more detailed description of this administrative database can be found elsewhere.[10]

Claims filed for work-related rape from 1980 to 1989 were identified by the Z16.2 nature of injury code for rape (Z16.2 code 600). Microfiche of claim records for these individuals were reviewed to validate three inclusion criteria: the occurrence of a rape, whether the claim was accepted as a work-related injury, and whether the establishment was covered by the state fund. Characteristics of the rape victims, types of establishments, and circumstances surrounding the incident were also recorded.

Employee hours worked at establishments covered by the state fund were obtained to estimate industry-specific rates. Age- and gender-specific proportions for each industry were determined from 1980 US census data for Washington State[11] for all persons over the age of 15 for whom a usual occupation was listed. Industry-specific rates of rape were calculated as rapes per 100 000 full-time equivalent female workers (2000 hours = 1 full-time equivalent female worker for 1 year).

Results

Eighty-one claims of women raped in the workplace during the study period were initially identified. Upon review, 18 claims filed were excluded, 7 because

Bruce H. Alexander and Marsha E. Wolf are with the Harborview Injury Prevention and Research Center, Seattle, Wash; they are also with the Department of Epidemiology, University of Washington, Seattle. Gary M. Franklin is with the Department of Environmental Health and the Department of Medicine (Neurology), University of Washington, Seattle, and the Department of Labor and Industries, Olympia, Wash.

Requests for reprints should be sent to Bruce H. Alexander, MS, Department of Epidemiology SC-36, University of Washington, Seattle, WA 98195.

This paper was accepted November 5, 1993.

TABLE 1—Age Distribution of Women Reporting Workplace Rape and Women Reporting an Occupation in the 1980 US Census of Washington State

Age, y	Rape Victims No.	Rape Victims %	Census, %[a]
<20	13	20.6	10.1
20–29	33	52.4	31.9
30–39	12	19.0	23.0
40–49	3	4.8	14.2
50+	2	3.2	20.7

[a]15 years of age and older.

they were not accepted by the Department of Labor and Industries as being work related, 3 because they involved other work-related injuries misclassified as rapes, and 8 because they involved businesses not covered by the state fund. The remaining 63 rapes were included in this analysis.

The rape victims tended to be younger than the population of women reporting a usual occupation (Table 1). More than half of the reported rapes occurred between the hours of 7 PM and 6 AM (Table 2). Fifty-three (85%) of the women were working alone when the rape occurred, a proportion that increased during the daytime hours. Fifty-four (86%) of the assailants were strangers, seven (11%) were coworkers, and two (3%) were patients in the institution at which the rape occurred. Robbery was a concurrent motive in 30% of the incidents, but in a smaller proportion during the day.

Twenty-two occupational groups were identified from the original claim forms (Table 3). Of note, five of the six food service workers were employed at fast food establishments, and four of the five bartenders worked in taverns. Two or more rapes were reported in 12 industries (Table 4), and one rape occurred in each of an additional 16 industries. The annual rate of work-related rape for all industries was 1.5 per 100 000 full-time equivalent female workers. The rate for industries reporting a rape was 4.4 per 100 000 full-time equivalent female workers. Taxi drivers had the highest rate of work-related rape, followed by job training and related services and veterinary services.

TABLE 2—Circumstances of Work-Related Rape Incidents Identified through Washington State Workers' Compensation Claims Data: 1980 to 1989

Circumstance	Subjects (n = 63), %	Time of Incident: 7 AM–6 PM No.	7 AM–6 PM %	7 PM–6 AM No.	7 PM–6 AM %
Hour of incident					
7 AM–12 PM	19.0	...	...	...	...
1 PM–6 PM	23.8	...	...	...	...
7 PM–12 AM	19.0	...	...	...	...
1 AM–6 AM	36.5	...	...	...	...
Unknown	1.6	...	...	...	...
On employer's property					
Yes	88.9	22	81.5	34	94.4
No	11.1	5	18.5	2	5.6
Working alone					
Yes	84.6	25	92.6	28	77.8
No	14.3	2	7.4	7	19.4
Unknown	1.6	0	0.0	1	2.8
Abducted					
Yes	15.9	4	14.8	6	16.7
No	84.1	23	85.2	30	83.3
Relationship to suspect					
Stranger	85.7	22	81.5	32	88.9
Coworker	11.1	5	18.5	2	5.6
Patient	3.2	0	0.0	2	5.6
Concurrent robbery					
Yes	30.2	6	22.2	13	36.1
No	69.8	21	77.8	23	63.9
Weapon involved					
Firearm	30.2	6	22.2	13	36.1
Knife	12.7	3	11.1	5	13.9
None	57.1	18	66.7	18	50.0

Discussion

The problem of intentional injuries in the workplace, especially nonfatal injuries, lacks a clear description. This dearth of information is, in large part, due to problems in identifying work-related assaults, particularly sexual assaults. Workers' compensation claims identify assaults only if they result in lost workdays or require medical treatment. Furthermore, only 16%[12] to 53%[13] of rapes are reported to authorities. Consequently, our study may substantially underestimate the incidence of work-related rape in this population. Nonetheless, some aspects of work-related rapes identified in this study are notable in their similarity to other reports of workplace violence and suggest that rapes occurring in the workplace are not random events.

This study identified occupations previously reported as involving high risk for rape in the workplace (convenience store clerks, property management personnel, and motel housekeepers)[7] and several additional occupations reporting more than one rape. These groups are comparable to occupational groups reported to have a greater risk of work-related homicide.[4–6,14]

Isolation of workers from fellow employees and the public appears to increase the risk of work-related rape. The rapes in this study frequently occurred between late evening and early morning. No data are available on the distribution of women working during these hours, yet the rapes occurred disproportionately during traditional nonworking hours. Furthermore, a large proportion of the rapes (85%) befell victims working alone at the time of the incident.

The calculated rates were unstable as a result of small numbers, and particular characteristics of the workplace environment cannot be linked to broad three-digit Standard Industrial Classification categories. Of interest, however, the rate for taxi drivers, an industry characterized by workers working in very isolated circumstances, was much higher than those for the other industrial groups.

TABLE 3—Occupations Reporting One or More Rapes between 1980 and 1989 in Washington State

Occupation[a]	Reported Rapes No. (n = 63)	%
Convenience store clerk	6	9.5
Food service	6	9.5
Retail sales	5	7.9
Nursing home and patient care	5	7.9
Bartender	5	7.9
Motel housekeeping and maintenance	5	7.9
Cleaning and janitorial	5	7.9
Secretary/receptionist	4	6.4
Bus/taxi driver	3	4.8
Hospital employee (pharmacy, technologists)	3	4.8
Gas station attendant	2	3.2
Security guard	2	3.2
Veterinary hospital assistant	2	3.2
Social services	2	3.2
Apartment management and maintenance	1	1.6
Dairy delivery	1	1.6
Electric company representative	1	1.6
Sales representative	1	1.6
Accountant	1	1.6
Librarian	1	1.6
Police dispatcher	1	1.6
Forest trail construction crew	1	1.6

[a]Identified from original claim report.

TABLE 4—Frequency of Reported Rapes and Estimated Rates per 100 000 Full-Time Equivalent (FTE) Female Workers, among Industrial Divisions Reporting Two or More Rapes between 1980 and 1989 in Washington State

SIC Code	Industry	Rapes	FTEs[a]	Rate[b]	95% Confidence Interval
412	Taxicabs	2	855	234.0	40.5, 939.2
833	Job training and related services[c]	3	11 342	26.5	6.8, 84.3
074	Veterinary services	2	8 838	22.6	3.9, 91.2
573	Radio, TV, and computer stores	2	13 089	15.2	2.6, 61.6
554	Gasoline service stations	2	19 174	10.4	1.8, 42.1
421	Trucking and courier services	2	20 116	9.9	1.7, 40.1
701	Hotels and motels	7	71 397	9.8	4.3, 21.2
599	Retail stores NEC	2	24 817	8.3	1.4, 32.5
541	Grocery stores	6	102 320	5.9	2.4, 13.5
806	Hospitals	3	59 205	5.1	1.3, 16.1
581	Eating and drinking establishments	12	437 746	2.7	1.4, 4.9
805	Nursing and personal care facilities	4	148 945	2.7	0.9, 7.4
	All industries reporting a rape	63	1 423 300	4.4	3.4, 5.7
	All industries	63	4 182 166	1.5	1.2, 1.9

Note. SIC = Standard Industrial Classification; NEC = not elsewhere classified.
[a]2000 worker hours per year.
[b]Rate of rapes per 100 000 full-time equivalent female workers employed by establishments covered by the Washington State fund.
[c]Includes job placement and counseling.

Intentional injury, including sexual assault in the workplace, is clearly an occupational safety problem. For more than a decade, public health researchers have recognized that safety standards to prevent assault in the workplace are needed to meet the changing problems of workplace safety.[2,14] Interventions have been suggested to reduce the risk of assault by creating a workplace environment that would deter an assailant.[1] Potential deterrents to assault should be more thoroughly examined for all occupations that involve a high risk of intentional injury. Clearly, workplace violence is linked to broader social issues that cannot be addressed from the standpoint of workplace safety alone. Nevertheless, to meet the goal of a safe work environment for all workers, the issue of violence in the workplace must be addressed. □

Acknowledgments

This research was supported by Harborview Injury Prevention and Research Center grant CCR 49-002570, Centers for Disease Control, and by the Accident and Medical Aid Funds of the State of Washington, Department of Labor and Industries. These research funds are targeted to reducing the incidence of and disability related to occupational injuries and illnesses. Bruce H. Alexander was supported by a training grant in environmental epidemiology and biostatistics from the National Institute for Environmental Health Sciences (T32ES07262).

This work was presented at the 121st Annual Meeting of the American Public Health Association, October 24–28, 1993, San Francisco, Calif (Session 3050, Abstract P.281).

We thank Drs Frederick P. Rivara and Peter Cummings for their review of and thoughtful comments on this manuscript.

References

1. Hales T, Seligman PJ, Newman SC, Timbrook CL. Occupational injuries due to violence. *J Occup Med.* 1988;30:483–487.
2. Baker SP, Samkoff JS, Fisher RS, Van Buren CB. Fatal occupational injuries. *JAMA.* 1982;248:692–697.
3. Bell CA, Stout NA, Bender TR, Conroy CS, Crouse WE, Myers JR. Fatal occupational injuries in the United States, 1980 through 1985. *JAMA.* 1990;263:3047–3050.
4. Davis H, Honchar PA, Suarez L. Fatal occupational injuries of women, Texas 1975–84. *Am J Public Health.* 1987;77:1524–1527.
5. Bell CA. Female homicides in United States workplaces, 1980–1985. *Am J Public Health.* 1991;81:729–732.
6. Davis H. Workplace homicides of Texas males. *Am J Public Health.* 1987;77:1290–1293.
7. Seligman PJ, Newman SC, Timbrook CL, Halperin WE. Sexual assault of women at work. *Am J Ind Med.* 1987;12:445–450.
8. *International Classification of Diseases, Ninth Revision, Clinical Modification.* 4th ed. Washington, DC: US Dept of Health and Human Services, Health Care Financing Administration; 1991. PHS publication 91-1260.
9. *Method of Recording Basic Facts Relating to the Nature and Occurrence of Work Injuries, Z16.2.* New York, NY: American National Standards Institute; 1962.
10. Franklin GM, Haug J, Heyer N, Checkoway H, Peck N. Occupational carpal tunnel syndrome in Washington State, 1986–1988. *Am J Public Health.* 1991;81:741–746.
11. *Census of Population and Housing, Washington State 1980. Public-Use Micro Data Sample C.* Washington, DC: US Bureau of the Census; 1983.
12. *Rape in America: A Report to the Nation.* Arlington, Va: National Victim Center and Crime Victims Research and Treatment Center; 1992.
13. *Female Victims of Violent Crime.* Washington, DC: US Dept of Justice; 1991. NCJ-126826.
14. Krauss JF. Homicide while at work: persons, industries, and occupations at high risk. *Am J Public Health.* 1987;77:1285–1289.

Preventing Occupational Illness and Injury: Nurse Practitioners as Primary Care Providers

Jane Lipscomb, RN, PhD, Barbara Burgel, RN, MS, Laura Wotowic McGill, RN, MS, and Paul Blanc, MD, MSPH

A B S T R A C T

Nurse practitioners with master's degrees were surveyed to assess the type and volume of occupational health services provided by primary care as compared with occupational health practitioners and the knowledge base in occupational health in these two groups. Thirty-six percent of 224 nonoccupational health nurse practitioners reported caseloads with 10% or more occupationally related chief complaints; 21% reported treating work-related injury or illness at least once per week. By contrast, a large percentage of nonoccupational health practitioners failed the knowledge-based exam. Large-scale prevention of occupational illness and injury warrants that primary care providers receive training in occupational health. (*Am J Public Health.* 1994;84: 643–645)

Introduction

Responsibility for most occupational health management falls to nonoccupational specialists.[1] The potential preventive impact of primary care health providers on this public health problem could be dramatic.[2] Nurse practitioners are an integral part of primary health care services in many practice settings. In contrast to primary care physician training, in which the need for occupational health education has received considerable attention, the lack of such training in nursing programs has been largely overlooked, and the amount of occupational health services provided by nurse practitioners has not been systematically evaluated.[1-7] To better understand the role of nurse practitioners in the recognition and prevention of occupational illness and injury, we surveyed occupational health services provided by primary care nurse practitioners and assessed their recognition of occupational illnesses and case management.

Methods

Study Site

We studied graduates of a large, university-based school of nursing that has offered graduate education in occupational health nursing since 1978, one of only seven master's-level programs in the country providing such specialized training. The occupational health nurse practitioner specialization is one of five 2-year nurse practitioner master's tracks within this school of nursing. The four others are adult, gerontology, family, and women's health. The occupational health nurse practitioner program includes a minimum of 14 academic quarter credits of didactic occupational health course work, other didactic community health nursing and primary care course work, and 540 practicum hours in primary care and occupational health settings. The adult nurse practitioner track (one of four nonoccupational health programs we surveyed) is closely affiliated with the occupational health nurse practitioner program, with students enrolled in a similar primary care curriculum but not required to engage in any occupational health specialty course work. In contrast, the gerontology nurse practitioner track shares only some of the primary care course work with the occupational health nurse practitioners and adult nurse practitioners, and students have less contact with the occupational health curriculum than the adult nurse practitioner students. Finally, the family health and women's health nurse practitioner tracks have separate curricula that do not overlap courses with the occupational health nurse practitioners.

Survey Methods

We surveyed the 509 nurse practitioner master's graduates for the years 1978 through 1989. The survey instrument covered current and past employment, the type and volume of occupational health services rendered in the past 6 months (including occupational health illness management, treatment of injuries, and health surveillance), and a knowledge assessment based on 10 multiple-choice clinical case scenarios. The text of the questionnaire is available upon request.

Statistical Analysis

We first compared proportions among occupational respondents and respondents from all other programs by

Jane Lipscomb, Barbara Burgel, and Laura Wotowic McGill are with the Department of Mental Health, Community, and Administrative Nursing, School of Nursing, University of California, San Francisco. Paul Blanc is with the Division of Occupational Medicine, University of California, San Francisco.

Requests for reprints should be sent to Jane Lipscomb, RN, PhD, Department of Mental Health, Community, and Administrative Nursing, School of Nursing, Box 0608, University of California, San Francisco, CA 94143-0608.

This paper was accepted August 10, 1993.

TABLE 1—Nurse Practitioners' Provision of Occupational Services, by Training Program of Graduates

Occupational Service	Nurse Practitioner Program: Occupational (n = 31), %	Nurse Practitioner Program: Nonoccupational (n = 223), %
Routine occupational history taking*	68	35
Frequent work-related chief complaints*	87	36
"Doctor's first report" completion*	81	32
Treatment of work injuries*	65	21
Treatment of illnesses	35	21
Provision of surveillance*	65	17

Note. Positive responses were defined as follows: occupational history taking among one third or more of clients, occupational chief complaint among 10% or more of clients, doctor's first report of occupational illness and injury ever filed on any client (legally required reporting procedure in California), and treatment of work injuries or illness and provision of surveillance once or more per week.
*$P < .001$ (occupational as compared with nonoccupational program graduates).

TABLE 2—Nurse Practitioners' Responses to 10 Clinical Case Scenarios on Occupational Health Test

Case Scenario Type	Occupational Program Graduates' No. of Correct Responses, Mean ± SD	Nonoccupational Program Graduates' No. of Correct Responses, Mean ± SD
Occupational disease identification*	3.9 ± 1.0	2.2 ± 1.2
Occupational case management*	4.1 ± 1.2	3.4 ± 1.5
Both case scenario types*	8.0 ± 1.7	5.6 ± 2.4

Note. Five clinical case scenarios were included in each category.
*$P < .05$ (occupational as compared with nonoccupational program graduates).

means of chi-square tests. We calculated Cronbach's alpha to evaluate the consistency of responses to the 10 case scenario questions. We compared the mean frequencies of correct responses by student's *t* test. A priori, we defined a "passing score" as 70% correct. We used logistic regression analysis to estimate the association between past training program and passing score, taking into account the effect of postgraduate occupational field experience in practice. The CRUNCH[8] and Epistat[9] statistical packages were used in conducting analyses.

Results

Of 509 questionnaires mailed, 23 (5%) were returned as nondeliverable. Three hundred thirty-seven questionnaires were completed, for an overall response rate of 69% of those deliverable. We excluded from further analysis 64 respondents not currently working in any clinical setting. We also excluded 18 additional respondents from the four nonoccupational health programs who were working in an occupational health rather than primary care setting. In total, we analyzed data from 254 respondents.

Of 31 occupational health program respondents, 22 (71%) were working in an occupational health setting (e.g., on-site occupational health service, occupational health specialty clinic). Four others were designated occupational health specialists in another clinical setting. Among 223 nonoccupational nurse practitioner program graduates, the breakdown by program was as follows: adult, 57 (26%); gerontology, 31 (14%); family, 74 (33%); and women's health, 61 (27%). Only 27 (12%) of the 223 nonoccupational respondents worked in inpatient settings. The remainder worked in the following settings: outpatient clinic of an inpatient center (n = 58), health maintenance organization (n = 51), private practice (n = 38), community clinic (n = 22), or "other" (n = 27).

Table 1 presents data for occupational health activities among occupational nurse practitioner program graduates as compared with all others. Occupational injury management included any of four specific services, illness management involved any of five services, and surveillance activities included any of seven specific activities. A greater proportion of occupational health graduates provided each type of occupational service; nonetheless, such activities were not infrequent for the nonoccupationally trained. For example, 21% of nonoccupational health graduate survey respondents reported providing at least one type of injury or illness management once or more per week. The most common specific services were for work-related stress, work-related back injury/chronic pain, work-related muscular strains other than back injury, and preplacement exams. Indeed, for occupational stress, the percentage of nonoccupational health nurse practitioners providing evaluations and/or treatment was higher than that among occupational health nurse practitioners (18% vs 13%). When defined by the less frequent criterion of treatment once per month rather than once per week, 50% among the nonoccupationally trained reported treating work-related musculoskeletal problems, and 20% reported treating work-related asthma. As shown in Table 1, 81 (36%) nonoccupational graduates reported 10% or more of their caseload as involving potentially occupationally related chief complaints. For this group, the overall mean frequency of occupationally related chief complaints was 9%.

We evaluated responses to the 10 case scenarios to assess the reliability of these questions in assessing a respondent's ability to correctly identify occupational illnesses and intervention. The Cronbach's alpha coefficient for the 10 cases was .73. The scenario test results for occupational health and nonoccupational health nurse practitioner graduates are presented in Table 2. Graduates of the occupational health program correctly identified a mean of 3.9 illness cases, in comparison with 2.2 by graduates of the nonoccupational health programs ($P < .001$). Of the 5 cases assessing intervention strategies, occupational health graduates correctly identified a mean of 4.1 interventions, as compared with 3.4 for nonoccupational health

graduates ($P < .012$). Defining a priori 70% as a passing test score, 13% of occupational health graduates failed, as compared with 60% of nonoccupational health graduates ($P < .05$). Taking into account year of training, frequency of occupational history taking, treatment of work-related chief complaints, and filing of "doctor's first reports," graduation from the occupational practitioner program study was the only factor significantly associated with a passing test score (odds ratio [OR] = 1.4, 95% confidence interval [CI] = 1.1, 1.7).

Discussion

Our study hypothesis was that nurse practitioners in primary care settings provide frequent occupational health services without the formal training to facilitate the prevention, diagnosis, and treatment of work-related illnesses and injuries. These survey data strongly support this hypothesis. Twenty percent of nonoccupational health graduates treated occupational injury and occupational illness at least weekly. Since we included gerontology and women's health specialty nurse practitioners, who are less likely to provide occupational health services, ours would be a conservative estimate of the types and frequency of services likely to be provided by general primary care nurse practitioners who see higher rates of patients in the labor force.

Despite these frequent primary occupational care responsibilities, only 4 in 10 nonoccupational health specialists achieved a passing score on a knowledge-based examination testing both diagnosis and case management in occupational health. The test appeared to have good internal and external validity.

Our findings are consistent with those from other studies of the potential importance of occupational health in primary care. In one study, 16% of patients of an outpatient practice with a predominantly "blue-collar" population reported that their health problem was related to work.[10] In another survey, 55 physicians in general practice reported that, on average, 3.5 patients per week had a work-related condition.[11] Assuming treatment of as many as 16 patients per day, this would approximate 4% of patients seen. Our findings of nonoccupational nurse practitioners reporting that 9% of their clients present with a potentially work-related chief complaint falls between that of patient self-reports and physician reports in these two published studies.

Our findings suggest that primary care nurse practitioners are on the "front line" for identifying and potentially preventing occupational illnesses and injuries. At the same time, this survey indicates that these health care providers are not provided the optimal training to identify, treat, or prevent this morbidity; they may be in the trenches, but they have not been supplied with enough ammunition. We believe that if primary care practitioners, including nurse practitioners, have the responsibility for diagnosing and managing occupational health problems, health care policymakers and educators have the duty to guarantee the training and resources needed to fulfill their duty. □

Acknowledgments

This study was supported in part by ERC Inc, University of Alabama, Birmingham (contract U60 CCU902886).

References

1. Institute of Medicine, Division of Health Promotion and Disease Prevention. *Role of the Primary Care Physician in Occupational Environmental Medicine.* Washington, DC: National Academy Press; 1988.
2. *Occupational Injuries and Illnesses in the US by Industry,* 1986. Washington, DC: US Dept of Labor; 1988. Bulletin 2308.
3. Felton JS. The occupational history: a neglected area in the clinical history. *J Fam Pract.* 1980;11:33–39.
4. Rosenstock L. Occupational medicine: too long neglected. *Ann Intern Med.* 1981;95:774–776.
5. Philipp R, Hughes AO. Hospital admissions and occupational histories. *Lancet.* 1983;1:129–130. Letter.
6. Goldman RH, Peters JM. The occupational and environmental health history. *JAMA.* 1981;246:2831–2836.
7. *Environmental Health: Occupational Health as a Component of Primary Health Care.* Copenhagen, Denmark: World Health Organization, Regional Office for Europe; 1986.
8. *CRUNCH Version 4 Statistical Package.* Oakland, Calif: Crunch Software Corp; 1991.
9. *Epistat Statistical Package.* Richardson, Tex: 1986.
10. Schwartz DA, Wakefield DS, Fieselmann JF, et al. The occupational history in primary care setting. *Am J Med.* 1991;90:315–319.
11. Merrill RN, Pransky G, Hathaway J, et al. Illness and the workplace: a study of physicians and employers. *J Fam Pract.* 1990;31:55–58.

Occupational Injury Deaths of 16- and 17-Year-Olds in the United States

Dawn N. Castillo, MPH, Deborah D. Landen, MD, MPH, and Larry A. Layne, MA

ABSTRACT

Data from the National Traumatic Occupational Fatalities surveillance system were used to analyze occupational injury deaths of civilian 16- and 17-year-olds during 1980 through 1989. There were 670 deaths; the rate was 5.11 per 100 000 full-time equivalent workers. The leading causes of death were incidents involving motor vehicles and machines, electrocution, and homicide. Workers 16 and 17 years old appear to be at greater risk than adults for occupational death by electrocution, suffocation, drowning, poisoning, and natural and environmental factors. Improved enforcement of and compliance with federal child labor laws, evaluation of the appropriateness of currently permitted activities, and education are encouraged. (*Am J Public Health.* 1994;84:646-649)

Introduction

Although several studies have examined nonfatal occupational injuries among youth,[1-8] only one published study has investigated occupational injury deaths of young people across the United States. Suruda and Halperin analyzed data from Occupational Safety and Health Administration (OSHA) fatality investigations between 1984 and 1987.[9] Of the 104 deaths investigated, 30% involved industrial vehicles and equipment, 17% were electrocutions, and 11% resulted from falls. Thirteen percent of the fatally injured youth were 15 years of age or younger.

Analysis of OSHA fatality investigations does not present a complete picture of occupational injury death in the United States, however. OSHA investigates only about 25% of work-related deaths, most of which are concentrated in the construction and manufacturing industries.[9] OSHA does not investigate homicides, most transportation incidents, and industries regulated by other federal agencies such as the Mine Safety and Health Administration and the Department of Transportation.

To obtain a more comprehensive picture of occupational injury deaths among youth, we analyzed data from the National Traumatic Occupational Fatalities surveillance system. This surveillance system is not limited by the industry or occupation in which the death occurred and includes all causes of occupational injury death. Because the minimum age included in the system is 16 years, the analysis was limited to 16- and 17-year-olds.

Methods

Data from the National Traumatic Occupational Fatalities surveillance system were used to analyze occupational injury deaths of 16- and 17-year-olds—and, for comparison, adults 18 years of age and older—for the years 1980 through 1989 (the most recent years for which data are available). This ongoing surveillance system, maintained by the National Institute for Occupational Safety and Health, includes information from death certificates for all work-related injury deaths reported by the 52 vital statistics reporting units in the United States (the 50 states, New York City, and the District of Columbia). To be included in the system, death certificates must meet the following criteria: the victim was at least 16 years of age; the death was attributed to an external cause (i.e., injury or poisoning), as designated by the *International Classification of Diseases*, ninth revision (ICD-9)[10]; and the certifier responded positively to the "injury at work" item on the death certificate. Information from the immediate, contributory, and underlying cause of death and injury description fields on the death certificate was used to code the external causes of death (ICD-9 codes E800 through E999). The industry division was coded from the "usual industry" field on the death certificate according to the *Standard Industrial Classification Manual, 1987*.[11]

For calculation of rates, data on hours worked were obtained from the Bureau of Labor Statistics' Current Population Survey, a monthly survey of US households selected from a probability sample representative of the civilian noninstitutional population.[12] For each age group, average hours worked were converted to full-time equivalent workers by multiplying the average number of hours worked per week by 52 weeks per year and dividing by 2000 (the average number of hours worked per year).

The authors are with the Division of Safety Research, National Institute for Occupational Safety and Health, Morgantown, WV.

Requests for reprints should be sent to Dawn N. Castillo, MPH, Division of Safety Research, National Institute for Occupational Safety and Health, 944 Chestnut Ridge Rd, MS 180, Morgantown, WV 26505.

This paper was accepted October 4, 1993.

Editor's Note. See related editorials by Pless (p 537) and Kaufer Christoffel (p 539) in this issue.

TABLE 1—Industry Division of 16- and 17-Year-Old Youth Who Died from Occupational Injury: United States, 1980 through 1989

Industry Division	Deaths
Not classified	316
Agriculture, forestry, and fishing	110
Construction	59
Services	56
Retail trade	42
Manufacturing	34
Transportation and public utilities	25
Mining	14
Wholesale trade	10
Public administration	3
Finance, insurance, and real estate	1
Total	670

Note. Because information on industry was available for only 53% of the fatally injured youth, interpretation should be restricted to identifying the minimum number of deaths within each industry division.

TABLE 2—Distribution and Rates of Occupational Injury Death, by Cause, for 16- and 17-Year-Olds and Adults 18 Years of Age and Older: United States, 1980 through 1989

	16- and 17-Year-Olds								
	Males		Females		Total		Adults		Rate Ratio[b]
	Deaths	Rate[a]	Deaths	Rate	Deaths	Rate	Deaths	Rate	
Motor vehicle related	149	2.05	13	0.22	162	1.24	14 113	1.41	0.88
Machine related	105	1.44	8	0.14	113	0.86	8388	0.83	1.04
Electrocution	78	1.07	2	ND	80	0.61	4384	0.43	1.42
Homicide	46	0.63	18	0.31	64	0.49	7512	0.75	0.65
Falls	37	0.51	1	ND	38	0.29	5941	0.59	0.49
Struck by falling object	31	0.43	0	ND	31	0.24	4103	0.41	0.59
Suffocation	29	0.40	0	ND	29	0.22	1175	0.12	1.83
Drowning	24	0.33	1	ND	25	0.19	869	0.09	2.11
Poisoning	20	0.28	0	ND	20	0.15	985	0.10	1.50
Natural and environmental	15	0.21	2	ND	17	0.13	758	0.08	1.63
Other	83	...	8	...	91	...	12 941	...	...
All causes	617	8.49	53	0.91	670	5.11	61 169	6.09	0.84

Note. ND = not determined because of instability of rates based on small numbers.
[a]Rate per 100 000 full-time equivalents (FTE). The employment figures used in calculation of rates for 16- and 17-year-olds were 7 270 705 male FTE, 5 837 621 female FTE, and 13 103 732 total FTE. The employment figure for adults was 1 000 000 000 FTE.
[b]16- and 17-year old occupational injury death rate/adult rate.

Rates of occupational injury death were calculated per 100 000 full-time equivalent workers.

Results

There were 673 occupational injury deaths to 16- and 17-year-olds for the 10-year period from 1980 through 1989. Military personnel accounted for 3 of these deaths (young people 17 years of age may enter the military with permission from their parents). These military cases were removed from further analysis because comparable employment data for military personnel were not available. The average annual rate of occupational injury death for 16- and 17-year-olds was 5.11 per 100 000 full-time equivalent workers; the rate for adults 18 years of age and older was 6.09 per 100 000.

Two hundred ninety-seven (44%) of the deaths occurred during the summer months (June through August). There were 143 deaths (21%) during September through November and approximately 100 deaths each during the winter and spring months.

There were 429 deaths involving 17-year-olds and 241 deaths involving 16-year-olds; 617 of the victims were male. Most of the victims (540) were White; 62 were Hispanic, 53 were Black, and 15 were of other or unknown race/ethnicity.

The rates of occupational injury death of 16- and 17-year-olds dropped from 7.16 per 100 000 full-time equivalent workers in 1982 to 3.94 per 100 000 in 1983 and have been stable since. The rate in 1989 was 3.67 per 100 000 full-time equivalent workers.

Occupation and industry information was not available for most cases. The "usual occupation" for 388 (58%) of the fatally injured youth was reported to be "student." Table 1 provides the frequency of deaths by industry division. Because industry was available for only 354 (53%) of the fatally injured youth, the interpretation of these data should be restricted to identifying the minimum number of deaths within each division.

The distribution of occupational injury deaths by cause and rates for each cause are presented in Table 2 for 16- and 17-year-olds and, for comparison, adults 18 years of age and older. The leading causes of occupational injury death for both age groups were incidents involving motor vehicles and machines. Although the overall risk of occupational injury death is lower for 16- and 17-year-olds than for adults, the risk for occupational injury death by electrocution, suffocation, drowning, poisoning, and natural and environmental causes appears to greater for young people than for adults (14 of the 17 youths who died from natural and environmental causes were struck by lightning).

Motor vehicles accounted for nearly 25% of the occupational injury deaths of 16- and 17-year-olds. Thirty-three percent (53) of the fatally injured youth were driving a motor vehicle, 17% (28) were passengers, and 20% (32) were pedestrians. Information included on the death certificate was insufficient to determine whether the fatally injured youth was a driver, passenger, or pedestrian in 28% (46) of the deaths involving motor vehicles.

Agricultural machinery accounted for 68% of the machine-related deaths; tractors accounted for 44% of the machine-related deaths. Of the 50 tractor-related deaths, at least 29 were caused by rollovers, and 4 were a result of the individual's being run over; power takeoffs contributed to at least 3 of the deaths.

Events leading to electrocution could not be determined for nearly 30% of the electrocutions. Contact with an energized power line was the most frequent cause, accounting for 55% (44) of electrocutions.

Discussion

Although the National Traumatic Occupational Fatalities surveillance system is the best available source of information on occupational fatalities in the United States for the period of this study, it has limitations. Identification of work-related injury deaths and appropriateness of occupation and industry information on the death certificate are the primary issues. In a review of 10 studies that used multiple sources to identify the universe of fatal work-related injuries, Stout and Bell reported that, on average, death certificates captured 81% of work-related injury deaths.[13] Thus, frequencies and rates presented in this paper should be considered conservative estimates. Studies have demonstrated agreement of 60% to 76% between "usual industry" and "usual occupation" entries on the death certificate and employment at the time of death.[14] On the basis of the high percentage of young people whose occupation was listed as "student" in this analysis, the correspondence appears to be poorer for youth. Jenkins et al. provide a thorough discussion of the methodology and limitations of the National Traumatic Occupational Fatalities surveillance system.[15]

Although occupational injury deaths declined in the early part of the decade, exposure to hazardous work environments remains a problem. The primary law designed to protect young people from adverse effects of employment is the Fair Labor Standards Act of 1938. The act sets minimum age standards for employment, limits hours and occupations in which 14- and 15-year-olds may work, defines occupations and tasks that are hazardous, and proscribes hazardous work for nonagricultural workers under 18 years of age[16] and agricultural workers under 16 years of age.[17] There are numerous exemptions to the act, the most notable being that children may work on the family farm at any age, at any time, and in any occupation, including work declared to be hazardous. In addition to the protection provided by federal law, all but four states—Colorado, Kansas, Maryland, and Mississippi—have child labor laws.[18]

The lack of detailed information available from death certificates precluded the enumeration of deaths in the National Traumatic Occupational Fatalities system resulting from activities permitted and not permitted under the Fair Labor Standards Act. Motor vehicle driving, which resulted in the deaths of at least 53 young people, is prohibited for nonagricultural workers under 18 years of age. Suruda and Halperin reported that 41% of occupational injury deaths of youth investigated by OSHA occurred while the child was engaged in work prohibited by federal child labor laws.[9] The small number of child labor inspectors and inadequate penalties have been cited as reasons for the persistence of child labor violations.[6,18] Increased coordination between OSHA and federal and state labor departments[9,19] and increased penalties sufficient to deter violations have been recommended as ways to improve enforcement and compliance with child labor laws.[19]

At least 110 of the fatally injured youth worked in the agriculture, forestry, and fishing industry. The Fair Labor Standards Act declares that operation of several types of agricultural machines, including tractors of more than 20-power-takeoff horsepower, is hazardous for those younger than 16 years of age. The appropriateness of hazardous agricultural work for young people deserves evaluation. Legislation recently enacted in Virginia that raised the minimum age from 16 to 18 for several hazardous agricultural occupations[20] should be evaluated. Formal training in the safe operation of farm equipment and the installation of protective devices, such as rollover protective structures, may also reduce the incidence of injury to children.

The appropriateness of youth employment in the construction industry should also be evaluated. The construction industry has the second highest rate of occupational injury death among workers of all ages[15] and contributed to the deaths of at least 59 young people.

In addition to regulation and enforcement, there is a role for education in preventing occupational injury of youth.[6] Young people, parents, teachers and school boards, physicians, and the business community must be educated about child labor laws and the risks of injury in the workplace. Youth and their parents require this information to make informed decisions about appropriate jobs for young people. Teachers, school boards, and physicians should consider these issues when dispensing work permits. The business community needs to be aware of its responsibilities under the law to protect youth in its employ and to understand the potential for tragedy related to noncompliance with existing laws.

This study of death certificate data has identified the leading causes of occupational injury death of 16- and 17-year-olds and examined differences in risk between youth and adults by external cause of death. Additional research is needed to understand the circumstances under which young people die from injuries incurred while working and to identify ways in which such deaths can be prevented. Epidemiologic data addressing the effect of training, job experience, supervision, job stressors, work permits, and child labor laws on the risk for occupational injury death of youth are needed. □

References

1. Brooks DR, Davis LK, Gallagher SS. Work-related injuries among Massachusetts children: a study based on emergency department data. *Am J Ind Med.* 1993;24:313–324.
2. Belville R, Pollack SH, Godbold JH, Landrigan PJ. Occupational injuries among working adolescents in New York State. *JAMA.* 1993;269:2754–2759.
3. Banco L, Lapidus G, Braddock M. Work-related injury among Connecticut minors. *Pediatrics.* 1992;89:957–960.
4. Heyer NJ, Franklin G, Rivara FP, Parker P, Haug JA. Occupational injuries among minors doing farm work in Washington State: 1986 to 1989. *Am J Public Health.* 1992;82:557–560.
5. Parker DL, Clay RL, Mandel JH, Gunderson P, Salkowicz L. Adolescent occupational injuries in Minnesota: a descriptive study. *Minn Med.* 1991;74:25–28.
6. Pollack SH, Landrigan PJ, Mallino DL. Child labor in 1990: prevalence and health hazards. *Annu Rev Public Health.* 1990;11:359–375.
7. Schober SE, Handke JL, Halperin WE, Moll MB, Thun MJ. Work-related injuries in minors. *Am J Ind Med.* 1988;14:585–595.
8. Coleman PJ, Sanderson LM. Surveillance of occupational injuries treated in hospital emergency rooms—United States, 1982. *MMWR.* 1983;32:31SS–37SS.
9. Suruda A, Halperin W. Work-related deaths in children. *Am J Ind Med.* 1991;19:739–745.
10. *International Classification of Diseases: Manual on the International Statistical Classification of Diseases, Injuries, and Causes of Death, Ninth Revision.* Geneva, Switzerland; World Health Organization; 1977.
11. *Standard Industrial Classification Manual, 1987.* Washington, DC: Office of Management and Budget; 1987.
12. *BLS Handbook of Methods.* Washington, DC: US Dept of Labor; 1988. Bulletin 2285.
13. Stout NS, Bell CA. Effectiveness of source documents for identifying fatal occupational injuries: a synthesis of studies. *Am J Public Health.* 1991;81:725–728.
14. Bell CA, Stout NA, Bender TR, Conroy

CS, Crouse WE, Myers JR. Fatal occupational injuries in the United States, 1980 through 1985. *JAMA*. 1990;263:3047–3050.
15. Jenkins EL, Kisner SM, Fosbroke DE, et al. *Fatal Injuries to Workers in the United States, 1980–1989: A Decade of Surveillance: National Profile*. Washington, DC: US Government Printing Office; 1993. DHHS publication NIOSH 93-108.
16. *Child Labor Requirements in Nonagricultural Occupations under the Fair Labor Standards Act* (Child Labor Bulletin no. 101). Washington, DC: US Dept of Labor, Employment Standards Administration, Wage and Hour Division; 1990. WH publication 1330.
17. *Child Labor Requirements in Agriculture under the Fair Labor Standards Act* (Child Labor Bulletin no. 102). Washington, DC: US Dept of Labor, Employment Standards Administration, Wage and Hour Division; 1990. WH publication 1295.
18. *Sacrificing America's Youth: The Problem of Child Labor and the Response of Government*. Chicago, Ill: National Safe Workplace Institute; 1992.
19. *Child Labor: Increases in Detected Child Labor Violations throughout the United States*. Washington, DC: US General Accounting Office; 1990. Publication GAO/HRD-90-116.
20. Nelson RR. State labor legislation enacted in 1992. *Monthly Labor Rev*. 1993; 116:35–49.

Head Injuries Incurred by Children and Young Adults during Informal Recreation

Susan P. Baker, MPH, Carolyn Fowler, PhD, Guohua Li, MD, DrPH, Margaret Warner, BS, and Andrew L. Dannenberg, MD, MPH

ABSTRACT

To examine injuries related to playground equipment, children's vehicles, roller skates, and skateboards, 1991 data on emergency room patients younger than 25 years in the Consumer Product Safety Commission's 91 surveillance hospitals were used. Head injury was the primary diagnosis for an estimated 58 480 patients, exceeding the total number of head injuries to bicyclists younger than 25 years. The head injury rate decreased with age. The large number and high rate of head injuries in children involved in a variety of recreational activities suggests the value of multipurpose helmets. (*Am J Public Health*. 1994;84:649–652)

Introduction

Sports and recreation account for more than 40% of all head injuries in children and adolescents, exclusive of those related to motor vehicles.[1] This study was undertaken to estimate the number of hospital-treated head injuries in children and young adults that occur in association with a variety of informal recreational activities. The activities studied involved playground equipment, children's vehicles, skateboards, and roller skates. Children's vehicles included tricycles, wagons, scooters, and wheeled riding toys.

The use of helmets, although increasing among bicyclists, is rare in other informal recreation. Of special interest was the possible value of multipurpose helmets that could be used in several activities.

Methods

The federal Consumer Product Safety Commission (CPSC) conducts active surveillance of injured patients treated at 91 hospital emergency rooms chosen to be nationally representative; this constitutes the National Electronic Injury Surveillance System. To develop national estimates, cases are weighted on the basis of the sampling fraction.[2] For activities in the present study, the average case sampled represented 53 patients.

For each activity listed in Table 1, CPSC provided a series of detailed tabulations for patients treated in 1991, including age by primary anatomic site of injury, nature of injury, sex, and disposition. Data on place of injury, missing for 37% of the patients, were not used in the analyses. Data on the other variables were complete for all patients. Head injuries were defined to include scalp lacerations, skull fractures, concussions, and other brain injuries; injuries to the face, mouth, eye, or ear were excluded.

Chi-square tests were used to assess age differences in the percentage of head injuries and differences in hospitalization rates between emergency room patients with head injuries and those without head injuries. For chi-square calculations, the actual number of patients in the sample was used rather than the estimated number in the general population.

For comparison with injuries incurred during these recreational activities, we analyzed injuries to bicyclists during 1991, using computerized data provided by CPSC.

Results

An estimated 464 000 persons younger than 25 years received emergency

The authors are with the Injury Prevention Center, The Johns Hopkins School of Public Health, Baltimore, Md.

Requests for reprints should be sent to Susan P. Baker, MPH, The Johns Hopkins Injury Prevention Center, 624 N Broadway, Baltimore, MD 21205.

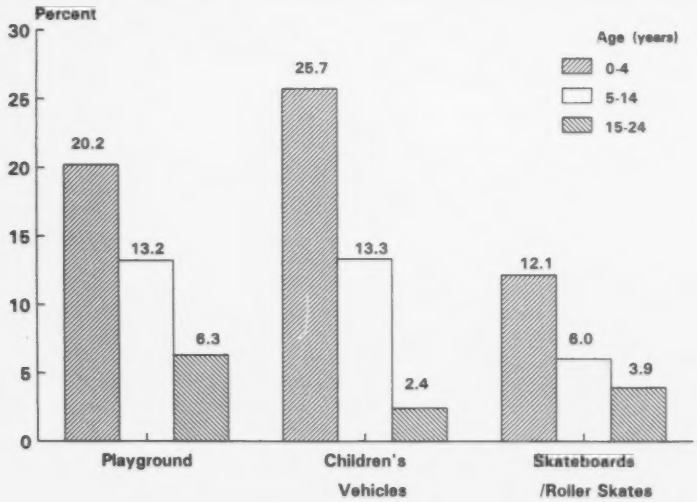
This paper was accepted November 12, 1993.

Editor's Note. See related editorial by Pless (p 537) in this issue.

TABLE 1—Estimated Head Injuries to Persons Younger than 25 Years Treated in US Emergency Rooms, by Equipment Associated with Injury and Patient's Age, 1991

Type of Equipment	Estimated No. of Head Injuries	Age of Patient					
		0–4 y		5–14 y		15–24 y	
		No.	%	No.	%	No.	%
Playground equipment	43 372	16 775	75	25 899	76	698	37
Slides	9309	5012	22	4297	13	0	0
Seesaws	1523	573	3	950	3	0	0
Monkey bars	8746	2360	10	6190	18	196	10
Swings	17 565	7409	33	9967	29	189	10
Trampolines	2487	259	1	1958	6	270	14
Other	2407	805	4	1559	5	43	2
Not specified	1335	357	2	978	3	0	0
Children's vehicles	6951	5029	22	1905	6	17	1
Tricycles	1899	1629	7	270	1	0	0
Wagons	2228	1565	7	663	2	0	0
Scooters	1207	507	2	683	2	17	1
Wheeled riding toys	1617	1328	6	289	1	0	0
Skateboards/roller skates	8157	684	3	6286	18	1187	62
Skateboards	4061	323	1	3118	9	620	33
Roller skates	4096	361	2	3168	9	567	30
Total	58 480	22 488	100	34 090	100	1902	100

Source. Data provided by the Consumer Product Safety Commission.



Source. Data provided by the Consumer Product Safety Commission.

FIGURE 1—Head injuries as a percentage of all injuries incurred during informal recreation, by age and activity, 1991.

room treatment during 1991 for injuries related to playground equipment, children's vehicles, skateboards, and roller skates. Motor vehicles were involved in 4% of all injuries related to children's vehicles and 1% of injuries related to skateboards and roller skates. The greatest difference between males and females was for skateboarding, with more than six times as many injuries in males as in females. Females received more roller skating injuries (1.4:1). For other activities, the male:female ratio ranged from 1:1 to 1.9:1.

Head Injuries

For an estimated 58 480 patients, 13% of the total, the head was the primary site of injury. Head injuries constituted 15% of all injuries associated with playground equipment, 20% of injuries from children's vehicles, and 6% of injuries from skateboards and roller skates. The proportion of injuries that involved the head decreased with age for each major category of activity; all were statistically significant at $P < .01$ (Figure 1). For informal recreational activities as a whole, head injuries constituted 21% of injuries to children younger than 5 years, 11% of injuries to children aged 5 through 14 years, and 5% of injuries to persons aged 15 through 24 years. Head injuries to bicyclists showed a similar pattern, decreasing with age as a proportion of all injuries.

For all ages combined, three fourths (74%) of the head injuries occurred in connection with playground equipment (Table 1). Swings, slides, and monkey bars were involved in 65% of head injuries from recreational activities for children younger than 5 years and 60% of those for children aged 5 through 14 years. Skateboarding and roller skating accounted for 62% of the head injuries from these activities for persons aged 15 through 24 years. The actual number of skateboard- and roller skate–related head injuries for persons aged 15 through 24 years was less than one fifth the number for the age group 5 through 14 years.

Children aged 5 through 14 years incurred 60% of all head injuries from activities involving playground equipment and 77% of head injuries involving skateboards and roller skates. Children younger than 5 years sustained 72% of the head injuries related to children's vehicles.

The total number of head injuries from the recreational activities studied (58 480) was even higher than the number of head injuries from bicycling (55 998) in the same age range. Head injuries from the recreational activities were more than twice as frequent as bicyclists' head injuries for children younger than 5 years and were 80% as frequent as head injuries from bicycling among children aged 5 through 14 years.

The head injury rate per 100 000 population for the selected activities was

highest for the age group 1 through 4 years. For bicycling, the head injury rate was highest for children aged 5 through 14 years (Figure 2).

Hospital admission was more than twice as common among patients with head injuries as among those without (5.7% vs 2.5%, $P < .001$) (Table 2). Head injuries associated with skateboards had an especially high proportion of admissions to hospitals (12.7%). Among bicyclists, 8.5% of those with head injuries were admitted, vs 1.7% of those without ($P < .001$).

Discussion

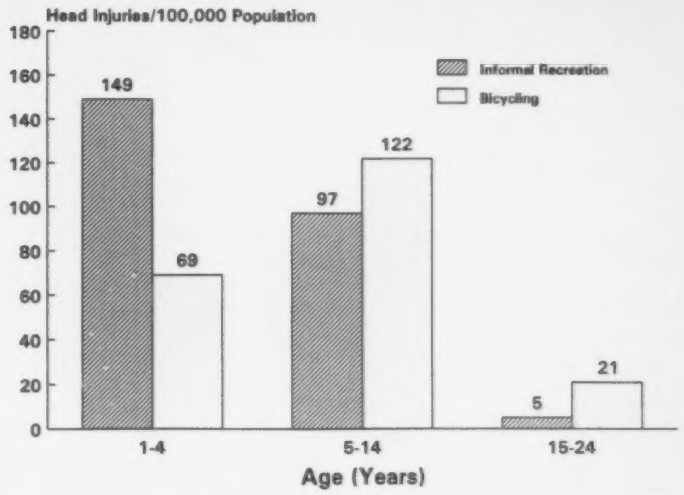
More than 58 000 head injuries occur annually in children and young adults in association with playground equipment, children's vehicles, skateboards, and roller skates. The percentage of injuries that involve the head is highest in children younger than 5 years, probably owing in part to the fact that the head is larger in proportion to the rest of the body in very young children. In addition, injury rates are highest in the youngest children. These facts, as well as the consequences of head injury, point to the need to take special precautions against head injury in children.

CPSC injury data are useful for providing national estimates but are limited by a lack of detail on injury severity; hospital admission is the only reliable severity indicator. Almost 6% of the head injuries studied resulted in hospital admission. In addition, some mild head injuries, including ones not resulting in hospital admission, have potentially severe long-term sequelae.[3]

Estimates of head injuries based on emergency room data are conservative because they exclude injured children who are treated outside hospitals or who do not seek medical attention. Such children may be less likely to have head injuries than those seen in emergency rooms, but data are lacking to clarify this point.

Population-based data on informal recreation are limited by the lack of exposure or denominator information that would identify recreational activities with a high risk of injury per unit of participation.[4] In contrast, for bicycling, one can use data from the Nationwide Personal Transportation Survey to calculate injury rates per million trips.[5] Even when high-risk activities cannot be singled out, however, the absolute number of injuries may suggest the need for environmental modifications[6] or protective headgear—for example, in the case of the use of swings, which were associated with 30% of all the head injuries.

Bicycle helmets have been shown to be effective in reducing head injuries[7,8] and are rapidly gaining in popularity. Because head injuries from other informal recreational activities are almost as common as head injuries in bicyclists aged 5 through 14 years and far more common in younger children, the poten-



Source. Data provided by the Consumer Product Safety Commission.

FIGURE 2—Incidence rates of head injury from bicycling and informal recreation, by age, 1991.

TABLE 2—Percentage of Emergency Room Patients Admitted to Hospitals, 1991, by Equipment Associated with Injury and Presence or Absence of Head Injury

Type of Equipment	Patients Hospitalized, %		*P*
	With Head Injuries	Without Head Injuries	
Playground equipment	4.5	2.7	<.01
Slides	6.4	2.3	
Seesaws	0.0	1.2	
Monkey bars	2.8	4.4	
Swings	5.7	1.9	
Trampolines	0.0	2.2	
Other	0.0	0.7	
Not specified	7.2	4.3	
Children's vehicles	1.7	1.0	>.57
Tricycles	6.1	0.7	
Wagons	0.0	1.1	
Scooters	0.0	1.2	
Wheeled riding toys	0.0	0.9	
Skateboards/roller skates	8.9	2.4	<.001
Skateboards	12.7	2.7	
Roller skates	5.0	2.3	
All types	5.7	2.5	<.001

Source. Data provided by the Consumer Product Safety Commission.

tial value of helmets for activities other than bicycling deserves emphasis.

Helmets are appropriate for a variety of activities at all ages, but are especially appropriate for children because of their high rates of head injury. Some children may participate in many recreational activities yet spend little time engaged in any single sport. Even when specialized helmets exist for a sport, a parent may not wish to invest in a single-purpose helmet for a child who is trying out an activity for the first time—yet inexperienced children are often at increased risk of injury.[9,10]

Multipurpose helmets are an appealing approach to the problem of head injury in informal recreational activities that lack specifically designed helmets. Rivara has urged the active promotion of helmets that can be used by children and adolescents in many sports.[1] Wilson et al. have pointed to the advantages of "a lightweight, inexpensive helmet that would be attractive to children and protective during a number of different recreational activities."[11]

Sports in which high-velocity impacts are common, such as skateboarding and in-line skating, may require helmets that differ from those giving the best protection in falls from playground equipment. It seems probable, however, that a helmet for young children could give good protection in more than one kind of activity. A standard for multipurpose helmets is currently being developed by the Snell Memorial Foundation (Channing Ewing, MD, MPH, telephone conversation, June 29, 1993).

The injury prevention community should be proactive in the area of informal recreational activities, which otherwise may not be subjected to the degree of scrutiny applied to team sports and other organized recreation. Concern should extend both to traditional activities, such as use of swings and tricycles, and to new activities. Recently we identified a surge in head injuries related to in-line skates,[12] illustrating the need for prompt recognition of hazards associated with new products. In any recreational activity where head injury prevention is a concern, early emphasis on helmet use may be more effective than trying to make a later change in existing behavior. □

Acknowledgments

This work was supported by The George Snively Research Foundation and Centers for Disease Control and Prevention grant R49/CCR302486.

This research would have been impossible without the substantial assistance of George Rutherford of the Division of Hazard and Injury Data Systems, Consumer Product Safety Commission.

References

1. Rivara FP. Childhood injuries, III: epidemiology of non-motor vehicle head trauma. *Dev Med Child Neurol.* 1984;26:81–87.
2. Consumer Product Safety Commission. The National Electronic Injury Surveillance System. *NEISS Data Highlights.* 1991.
3. Katz RT, Deluca J. Sequelae of minor traumatic brain injury. *Am Fam Physician.* 1992;46:1491–1498.
4. Runyan C, Gerken EA. Epidemiology and prevention of adolescent injury. *JAMA.* 1989;262:2273–2279.
5. Baker SP, Li G, Fowler C, Dannenberg AL. *Injuries to Bicyclists: A National Perspective.* A report to the Snell Memorial Foundation. Baltimore, Md: The Johns Hopkins Injury Prevention Center; 1993.
6. *A Handbook for Public Playground Safety, Vol. 2. Technical Guidelines for Equipment and Surfacing.* Bethesda, Md: Consumer Product Safety Commission; 1981.
7. Brown B, Farley C. The pertinence of promoting the use of bicycle helmets for 8- to 12-year-old children. *Chron Dis Can.* 1989;10:92–94.
8. Thompson RS, Rivara FP, Thompson DC. A case–control study of the effectiveness of bicycle safety helmets. *N Engl J Med.* 1989;320:1361–1367.
9. Cass DT, Ross F. Skateboard injuries. *Med J Aust.* 1990;153:140–144.
10. Inkelis SH, Stroberg AJ, Keller EL, Christenson PD. Roller skating injuries in children. *Pediatr Emerg Care.* 1988;4:127–132.
11. Wilson MH, Baker SP, Teret SP, Shock S, Garbarino J. *Saving Children: A Guide to Injury Prevention.* New York, NY: Oxford University Press; 1991:135.
12. Baker SP, Fowler C, Li G, Warner M. *Head Injuries in Non-Motorized Informal Recreation.* A report to the George Snively Research Foundation. Baltimore, Md: Johns Hopkins Injury Prevention Center; 1992.

A Cost–Benefit Analysis of Legislation for Bicycle Safety Helmets in Israel

Gary M. Ginsberg, DrPH, MSc, and Don S. Silverberg, MD

ABSTRACT

Legislation requiring bicyclists to wear helmets in Israel will, over a helmet's 5-year duration (assuming 85% compliancy, 83.2% helmet efficiency for morbidity, and 70% helmet efficiency for mortality), save approximately 57 lives and result in approximately 2544 fewer hospitalizations; 13 355 and 26 634 fewer emergency room and ambulatory visits, respectively; and 832 and 115 fewer short-term and long-term rehabilitation cases, respectively. Total benefits ($60.7 million) from reductions in health service use ($44.2 million), work absences ($7.5 million), and mortality ($8.9 million) would exceed program costs ($20.1 million), resulting in a benefit–cost ratio of 3.01:1. (*Am J Public Health.* 1994;84:653–656)

Introduction

Almost all bicycle accident mortality results from head and brain damage,[1–5] and nonfatal head injuries sometimes result in victims requiring expensive, lifelong care for their disabilities.[6–11] Between 70.9% and 95.5% of head injuries and about 70% of fatalities involving head injuries could be prevented through the use of crash helmets.[12–21] Fewer than 2% of Israel's estimated 833 000 cyclists[22] currently use such helmets.[11,23]

During the period from 1967 to 1991,[24] accidents to bicyclists accounted for annual mortality, hospitalization,[23] and emergency room visitation rates of 5.02, 228.5, and 2262.4 per million, respectively,[24] after adjustment for the estimated 88.9% (95% confidence interval [CI] = 87.7%, 90.1%) underreporting of emergency room cases.[23,25] Hospitalization rates were stable throughout this period, while mortality rates decreased between 1975 and 1980 and then stabilized. Persons under 20 years old accounted for 57.0% of hospitalizations, and older individuals had higher case fatality rates[24] (Table 1).

This paper estimates the costs and benefits from expected reductions in mortality and morbidity if mandatory bicycle helmet use legislation were enacted in Israel.

Methods

Direct (i.e., including only health service benefits) and total (direct, work, and mortality) or societal benefit–cost ratios based on the parameters in Table 2 were calculated on a spreadsheet; a 5% per annum discount rate was used. Lifetime costs of premature mortality, institutional care, and special education were discounted and attributed to the year in which the accident occurred.

The 85% compliancy rate (see Table 2) is based on Israelis' 90% compliancy with car seat belt laws,[26] only 6% less than the 96% achieved for bicycle helmet use after legislation in Melbourne, Australia.[21]

The gross national product per capita method of valuing life,[27] which assigns an equal value to each member of society equivalent to the annual per-capita gross national product of the population ($10 728 in 1990), was used to estimate mortality costs. These costs were based on discounting the value of the deceased's expected years of life lost.

Results

In 1994, bicycle-related injuries will account for approximately 26.5 deaths, 1203 hospitalizations, and 11 913 persons requiring emergency room treatment.

The cost of crash helmets (current retail price: $23.40) for Israel's 833 000 cyclists would be $19.5 million. An additional $607 000 would be required over 5 years for health education purposes.

Legislation for bicycle helmet use in Israel would, over a helmet's 5-year duration,[28] prevent approximately 57 deaths. Also, it would result in 2544 fewer hospitalizations, 13 355 fewer emergency room visits, 26 634 fewer ambulatory visits, and 832 and 115 fewer short- and long-term rehabilitation cases (including 45 children needing special education), respectively (Table 3).

The direct benefit ($44.20 million) to cost ($20.14 million) ratio is 2.19:1. When benefits from reductions in work absences ($7.55 million) are included, the benefit–cost ratio rises to 2.57:1. After adding benefits from reductions in mortality ($8.94 million), the total benefit–cost ratio to society rises to 3.01:1 (Table 3).

Sensitivity analyses available upon request from the authors show the benefit–cost ratio to be sensitive to

Gary M. Ginsberg is with the Department of Data Analysis, Division of Computing and Information, Ministry of Health, Jerusalem, Israel. Don S. Silverberg is with the Department of Public Health, Tel Aviv, Israel.

Requests for reprints should be sent to Gary M. Ginsberg, DrPH, MSc, Reuven 25/2, Baka, Jerusalem, 93510, Israel.

This paper was accepted July 14, 1993.

TABLE 1—Deaths, Hospitalizations, and Light Injuries from Bicycle Injuries, by Age: 1979 through 1990, Israel

Age, y	Deaths (n = 184), %	Hospitalizations (n = 2177), %	Light Injuries (n = 9071), %
0–4	0.5	1.0	0.9
5–9	9.2	12.6	8.0
10–14	14.7	25.6	22.1
15–19	12.0	17.8	19.7
20–24	2.7	6.3	7.6
25–34	8.2	7.1	10.5
35–44	7.6	5.5	7.5
45–54	10.9	7.4	7.2
55–64	12.5	7.1	7.9
65+	21.7	9.5	8.7
Total	100.0	100.0	100.0

TABLE 2—Parameters Used in Cost–Benefit Analysis

	Parameter	Source
Israeli population (1992)	5 266 000	(Reference no). 42
Per capita gross national product, $	10 728	43
Average annual wage costs, $	14 338	43
Average income of person in bicycle accident, $	5 519	Calculations on ref 24
Unemployment (1992), %	12.0	42
Male accident victims, %	87.0	24
Helmet cost, $	23.40	Retailers
Health education costs (5 years), $	607 000	Ministry of Health
Bicycle ownership rate, %	8.55	22
Riders per bike	1.85	22
Duration of helmet life, y	5	28
Initial compliancy, %	2.00	6
Final compliancy, %	85.00	5% less than refs 14, 21
Mortality		
Deaths due to head injuries, %	74.05	Midpoint range, refs 1–4, 14, 20, 44–46
Helmet efficiency, %	70.00	1
Dead on arrival at hospital, %	67.06	1
Average survival time, d	11	1
Cost of intensive care per day, $	1 057	47, 48
Average age at death, y	38.2	24
Life expectancy at 38.2 yrs.	38.3	43
Value of life at 38.2 years, $	181 449	Gross national product method
Value of life per deceased, $	170 519	Human capital method
Hospitalizations		
Helmet efficiency, %	83.20	Midpoint range, refs 12–16, 21, 49, 50
Serious head injuries, %	70.00	15, 26, 44, 46, 51–54
Average length of stay, d	9.7	Neurology department, ref 10
Cost of neurology department per day, $	446	55
Head injuries with other diagnoses, %	12.50	6
Average length of stay, d	4.5	6
Internal/orthopedic department cost per day, $	238	56
Emergency room		
Helmet efficiency, %	83.20	Midpoint range, refs 12–16, 21, 49, 50
Facial trauma alone, %	20.00	17, 57
Mild head trauma, %	24.36	4, 51–53, 58, 59
Severe head trauma, %	8.11	4, 51–53, 58, 59
Average cost per visit, $	117.50	Half of full day's hospitalization
Males, %	87.0	24

(continued)

changes in the discount rate, duration of helmet life, compliancy rate, and helmet effectiveness and cost.

Direct benefit breakeven points (where direct benefits are equal to costs) are met if compliancy falls below 40%, helmets last only 25 months, the helmet effectiveness rate is 37.8%, helmets cost $52.29, or there are more than 4.15 riders per bike. Breakeven points to society are met if compliancy falls below 29.6%, helmets last only 18 months, the helmet effectiveness rate is 17.8%, helmets cost $72.08, or there are more than 5.68 riders per bike.

Changes in the percentage of emergency room visits, unemployment, head injury death, and bike ownership rates, as well as use of the human capital method to value life, had little effect on the benefit to cost ratios.

Discussion

If compulsory bicycle crash helmet legislation were enacted in Israel, considerable reductions in mortality, morbidity, and residual disablement would occur, resulting in net savings of $43.3 million to society over a 5-year period. In almost all of the sensitivity analysis scenarios, both direct and total benefits exceeded program costs.

The benefit–cost ratios are biased downward since our conservative model did not consider reduced pain, worry, grief, work losses for ambulatory visits, or even time off from housework as a result of bicycle injuries. Nor did we consider the intangible benefit of the lessening of anxiety concerning crashes by cyclists or by their friends and relatives. Our calculations excluded revenues from possible fines as a consequence of the legislation, since payments of fines represent simply transfer payments that result in neither a net gain nor a net loss to society.

Another conservative assumption was that helmets would not reduce severe facial injuries, since recent experience revealed a reduction in severe upper facial injuries of approximately 73% as a result of helmet use.[17] Including benefits from decreased upper facial injuries only slightly increased the direct and total benefit–cost ratios to 2.27:1 and 3.13:1, respectively.

Those who object to bicycle helmet legislation argue that legislation infringes on the freedom of the individual. However, there is a legal precedent created by the fact that already-estab-

lished legislation for motorcycle helmets in Israel has resulted in reduced morbidity and mortality.[24] Studies conducted in the United States have shown huge health service cost savings as a result of motorcycle crash helmet legislation.[29–34]

Health education campaigns alone (even with bicycle helmet subsidies) have achieved only low compliancy rates ranging from 7.2% to 32.2%.[14,21,35–39] However, when legislation was added to a health education campaign in Victoria, Australia, compliancy rates rose from 30% to 90% in under a year.[14,21] Similar legislation for children under 16 years of age in Howard County, Maryland, increased compliancy to 47% relative to 19% from a community education program alone in a neighboring county.[40]

The imposition of legislation, backed by the threat of fines, is a powerful and preferred means of raising helmet wearing compliancy. The monetary benefits are sufficiently large to warrant the government ministries and/or the major health insurance funds to consider subsidizing[41] the costs of providing helmets to Israeli bicyclists. □

TABLE 2—Parameters Used in Cost–Benefit Analysis (*Continued*)

	Parameter	Source
Emergency room *(continued)*		
Workdays lost	5.56	23
Average wages lost per cohort member per day, $	21.39	43
Ambulatory visits		
Visits per emergency room survivor	2.00	Panel of doctors
Time of visit, min	15	Panel of doctors
Physician cost per year, $	32 640	60
Physician cost per hour, $	16	60
Short-term disabilities		
Emergency room visits with temporary cognitive disability, %	0.68	2, 61
Cost of rehabilitation care per day, $	7 250	55
Average length of stay, d	89.1	62
Long-term disabilities		
Emergency room visits with permanent cognitive disability, %	0.0028	2, 10, 11
Average age, y	27.0	51
Cost of institutional care per day, $	40.45	Ministry of Labor and Social Welfare
Average length of stay, y	40	51
Lifetime care costs, $	253 000	Discounted
Employable, %	15	Panel of doctors
Working years lost	38	30
Special education		
Extra cost per year, $	7 841	Ref 30, Ministry of Education
Average years needed	5.9	51

TABLE 3—Benefits of Crash Helmet Legislation for Bicyclists for 5 Years[a]

	Before	After	Reduction	Benefits, $	Benefit to Cost Ratio
Deaths	132	75	57	188 773	
Hospitalizations	6 016	3 471	2 544	10 549 317	
Emergency room visits	59 562	46 206	13 355	1 358 673	
Ambulatory care	118 946	92 312	26 634	92 249	
Rehabilitation	1 205	373	832	5 223 610	
Total health service benefits				17 412 622	0.86
Long-term care	167	52	115	25 263 243	
Special education	65.4	20.2	45	1 527 131	
Total direct care benefits				44 202 996	2.19
Mortality benefits				8 939 979	0.44
Work benefits[b]				7 545 779	0.37
Total benefits to society				60 688 755	3.01
Costs				20 143 985	
Total resource savings				43 276 133	

[a]Based on 5.0% discount rate, 85% final compliancy rate, 2% baseline compliancy, 12% unemployment rate, 8.55% bicycle ownership rate, 1.85 riders per bike, helmet effectiveness rates of 83.2% (injuries) and 70% (deaths), 5-year duration of helmet costing $23.40, 70% of deaths and 32.47% of emergency room cases from head injuries, 11.07% of emergency room visits reported, and 0% reduction in facial injuries.

[b]$6 198 569 from reduction in permanent disabilities and $1 347 210 from reduction in emergency room visits.

References

1. Fife D, Davis J, Tate L, Wells JK, Mohan D, Williams A. Fatal injuries to bicyclists: the experience in Dade County, Florida. *J Trauma.* 1983;23:745–755.
2. McDermott FT, Klug GL. Head injury predominance: pedal cyclists vs motor cyclists. *Med J Aust.* 1985;143:232–234.
3. Nixon J, Clacher R, Pearn J, Corcoran A. Bicycle accidents in childhood. *Br Med J.* 1987;294:1267–1269.
4. Sacks JJ, Holmgreen P, Smith SM, Sosin DM. Bicycle-associated head injuries and deaths in the United States from 1984 through 1988. *JAMA.* 1991;266:3016–3018.
5. Schmidt P, Haarhoff K, Bonte W. Fatal bicycle accidents in Dusseldorf 1980–1989. Analysis of autopsy cases. *J Traffic Med.* 1991;19:113–118.
6. Rimel RW, Giordani B, Barth JT, Boll TJ, Jane JA. Disability caused by minor head injury. *Neurosurgery.* 1981;9:221–228.
7. Lundar T, Nestvold K. Pediatric head injuries caused by traffic accidents: a prospective study with 5-year follow-up. *Child Nerv Syst.* 1985;1:24–28.
8. Kraus JF, Fife D, Couray C. Incidence, severity and outcomes of brain injuries involving bicycles. *Am J Public Health.* 1987;77:76–78.
9. Nakayama DK, Gardner MJ, Rogers KD. Disability from bicycle-related injuries in children. *J Trauma.* 1990;30:1390–1394.
10. Levi L, Lynn S, Gilbourt I, Feinsod M. Head and brain trauma as a result of bicycle accidents [in Hebrew]. *Harefuah.* 1990;118:575–576.
11. Silverberg D, Meer A, Silvinger E, Gross M, Feldman S. Head injuries after serious bicycle accidents. *Eur J Epidemiol.* 1992;8:826–831.
12. Dorch MM, Woodward AJ, Somers RL. Do bicycle safety helmets reduce severity of head injury in real crashes? *Accident Analysis Prev.* 1987;19:183–190.

13. Wasserman RC, Waller JA, Monty MJ, Emery AB, Robinson DR. Bicyclists, helmets and head injuries: a rider-based study of helmet use and effectiveness. *Am J Public Health.* 1988;78:1220–1221.
14. Wood T, Milne P. Head injuries to pedal cyclists and the promotion of helmet use in Victoria, Australia. *Accident Analysis Prev.* 1988;20:177–185.
15. Thompson RS, Rivara FP, Thompson DC. A case control study of bicycle safety helmets. *N Engl J Med.* 1989;320:1361–1367.
16. Wasserman RC, Buccini RV. Helmet protection from head injuries among recreational bicyclists. *Am J Sports Med.* 1990;18:96–97.
17. Thompson DC, Thompson RS, Rivara FP, Wolf ME. A case control study of the effectiveness of bicycle safety helmets in preventing facial injury. *Am J Public Health.* 1990;80:1471–1474.
18. Runyan CW, Earp JAL, Reese RP. Helmet use among competitive cyclists. *Am J Prev Med.* 1991;7:232–236.
19. Williams M. The protective performance of bicyclist's helmets in accidents. *Accident Analysis Prev.* 1991;23:119–131.
20. Spaite DW, Murphy M, Criss EA, Valenzuela TD, Meislin HW. A prospective analysis of injury severity among helmeted and nonhelmeted bicyclists involved in collisions with motor vehicles. *J Trauma.* 1991;31:1510–1516.
21. Vulcan AP, Cameron MH, Watson WL. Mandatory bicycle helmet use, experience in Victoria, Australia. *World J Surg.* 1992;16:389–397.
22. Katz A. Some characteristics of bicycle travel and accidents in towns. *Transportation Res Rec Transportation Res Board.* 1978;683:25–33.
23. Avitsur M. *Traffic Injuries, Causes, Methods of Transportation and Results.* Jerusalem, Israel: Department of Community Medicine, School of Public Health and Social Medicine, Hebrew University; 1992.
24. *Road Accidents with Casualties—1967 to 1991.* Jerusalem, Israel: Central Bureau of Statistics; 1968–1992.
25. Gofin R, Palti H, Adler B, Edet E. Childhood injuries: a population based study of emergency room visits in Jerusalem. *Paediatr Perinatol Epidemiol.* 1989;3: 174–178.
26. Zedar A, Mokous D, Hakert S, et al. *Seat Belt Use in Cities, Analysis of Utilization and Evaluation of Safety.* Haifa, Israel: Transportation Research Institute, Technion-Israel Institute of Technology; 1989. Research Report No. 145.
27. Ginsberg GM. Cost-effectiveness analysis, cost-benefit analysis and the value of life. In: Laaser U, Roccella EJ, Rosenfeld JB, Wenzel H, eds. *Costs and Benefits in Health Care and Prevention.* Berlin: Springer-Verlag; 1990:6–19.
28. Wilson MH, Baker SP, Teret SP, Shock S, Garbarino J. *Saving Children—A Guide to Injury Prevention.* New York, NY: Oxford University Press; 1991:78.
29. Shankar BS, Ramzy AI, Soderstrom CA, Discinger PC, Clark CC. Helmet use, patterns of injury, medical outcome, and costs among motorcycle drivers in Maryland. *Accident Analysis Prev.* 1992;24:385–396.
30. Offner PJ, Rivara FP, Maier RV. The impact of motorcycle helmet use. *J. Trauma.* 1992;32:636–641.
31. Braddock M, Schwartz R, Lapidus G, Banco L, Jacobs L. A population-based study of motorcycle injury and costs. *Ann Emerg Med.* 1992;21:273–278.
32. Muelleman RL, Mlinek EJ, Collicott PE. Motorcycle crash injuries and costs: effect of a reenacted comprehensive helmet use law. *Ann Emerg Med.* 1992;21:266–272.
33. Kelly P, Sanson T, Strange G, Orsay E. A prospective study of the impact of helmet usage on motorcycle trauma. *Ann Emerg Med.* 1991;20:852–856.
34. McSwain NE, Belles A. Motorcycle helmets—medical costs and the law. *J Trauma.* 1990;30:1189–1197.
35. Pendergast RA, Seymore-Ashworth C, Durant RH, Litaker M. Correlates of children's bicycle helmet use and short-term failure of school-level interventions. *Pediatrics.* 1992;90:354–358.
36. Weiss BD. Trends in bicycle helmet use by children: 1985 to 1990. *Pediatrics.* 1992;89:78–80.
37. Cushman R, James W, Waclawik H. Physicians promoting bicycle helmets for children: a randomized trial. *Am J Public Health.* 1991;81:1044–1046.
38. Cushman R, Down J, MacMillan N, Waclawik H. Helmet promotion in the emergency room following a bicycle injury: a randomized trial. *Pediatrics.* 1991; 88:43–47.
39. Bergman AB, Rivara FP, Richards DO, Rogers LW. The Seattle children's bicycle helmet campaign. *Am J Dis Child.* 1990;144:727–731.
40. Cote TR, Sacks JJ, Lambert-Huber DA, et al. Bicycle use among Maryland children: effect of legislation and education. *Pediatrics.* 1992;89:1216–1220.
41. Capron AM. Bicycle safety helmets. *N Engl J Med.* 1989;321:1194. Letter.
42. *Monthly Bulletin of Statistics.* Jerusalem, Israel: Central Bureau of Statistics; December 1992;43:12.
43. *Statistical Abstract of Israel 1991.* Jerusalem, Israel: Central Bureau of Statistics; 1991.
44. Reynolds R, Cohen D. Children, bicycles and head injuries. *Med J Aust.* 1987;146: 615–619.
45. Guichon DMP, Myles ST. Bicycle injuries: one year sample in Calgary. *J Trauma.* 1975;15:504–506.
46. Bouvier R. Bicycle accidents in childhood: social and preventive aspects. *Aust Fam Physician.* 1984;13:287–289.
47. Ginsberg GM, Shouval D. Cost-benefit analysis of a nationwide neonatal inoculation programme against hepatitis B in an area of intermediate endemicity. *J Epidemiol Community Health.* 1992;46:587–594.
48. *Hospitalization Prices from 1.1.92.* Jerusalem, Israel: Ministry of Health; 1991.
49. Thompson RS, Rivara FP, Thompson DC. Bicycle safety helmets. *N Engl J Med.* 1989;321:1195–1196. Letter.
50. Thompson DC, Thompson RS, Rivara FP. Incidence of bicycle-related injuries in a defined population. *Am J Public Health.* 1990;80:1388–1389.
51. Siebst SM, Alexander D, Ruddy R. Bicycle related injuries. *Am J Dis Child.* 1987;141:140–144.
52. Bjornstig V, Naslund K. Pedal cycling accidents—mechanisms and consequences. *Acta Chir Scand.* 1984;150:350–359.
53. McKenna PJ, Welsh DJ, Martin LW. Pediatric bicycle trauma. *J Trauma.* 1991; 31:392–394.
54. Arinson CJ, Pollard CW. Child cyclist injuries: a prospective study. *Med J Aust.* 1986;144:144–146.
55. *Departmental Costing of Ezrat Nashim Hospital 1987.* Jerusalem, Israel: Health Economics Consultants; 1988.
56. Ginsberg GM, Penchas S, Israeli A. Alternative methods of hospital remuneration in Israel. *Isr J Med Sci.* 1991;27: 583–589.
57. Hodgson VR. *Impact, Skid and Retention Tests on a Representative Group of Bicycle Helmets to Determine Their Head-Neck Protective Characteristics.* Detroit, Mich: Wayne State University, Department of Neurosurgery; 1990.
58. Centers for Disease Control. Bicycle-related injuries; data from the National Electronic Injury Surveillance System. *MMWR.* 1987;86:269–271.
59. *Accident Facts, National Safety Council.* Chicago, Ill: National Safety Council; 1982:45–91.
60. Ginsberg GM, Tulchinsky TH. Costs and benefits of a second measles inoculation of children in Israel, the West Bank and Gaza. *J Epidemiol Community Health.* 1990;44:274–280.
61. Kraus JF, Rock A, Hemyar P. Brain injuries among infants, children, adolescents and young adults. *Am J Dis Child.* 1990;144:684–691.
62. *Hospitals and Day Care Units in Israel. Statistical Information on Patient Movements by Institution and Department 1990.* Jerusalem, Israel: Ministry of Health, 1991.

Adolescent Occupational Injuries Requiring Hospital Emergency Department Treatment: A Nationally Representative Sample

Larry A. Layne, MA, Dawn N. Castillo, MPH, Nancy Stout, EdD, and Patricia Cutlip

ABSTRACT

Data from a nationally representative sample of emergency departments for the 6-month period July through December 1992 were used to examine nonfatal occupational injuries sustained by adolescents aged 14 through 17 years. There were 679 occupational injuries, corresponding to an estimated 37 405 injuries nationwide. Males constituted 65.8% of the injury victims. The injury rate for males was 7.0 per 100 full-time employees, compared with 4.4 for females. Lacerations to the hand or finger accounted for 25.6% of all injuries. The majority of injuries occurred in retail trades (53.7%), which also had the highest rate (6.3 per 100 full-time employees). Seventy-one percent of the injuries in retail trade occurred in eating and drinking establishments. (*Am J Public Health.* 1994;84:657–660)

Introduction

Among affluent countries, the United States has the largest proportion of its children in the workforce.[1] Although the psychosocial effects of youth employment are controversial,[2–4] the potential for physical injury must be recognized. Castillo et al. report an occupational fatality rate of 5.1 per 100 000 full-time employees for 16- and 17-year-olds.[5] Many fatally injured youths were involved in activities prohibited by federal child labor laws.[5,6]

Several studies have examined nonfatal occupational injuries among adolescents.[7–14] The only published study that reported rates for the entire United States used 1982 data from the National Electronic Injury Surveillance System (NEISS). The rates for 16- and 17-year-olds, 8.2 per 100 full-time males and 3.0 per 100 full-time females, were exceeded only by the rates for 18- and 19-year-olds.[15]

In this study we use NEISS data to describe nonfatal occupational injuries sustained by youths aged 14 through 17 years in the latter 6 months of 1992 that required hospital emergency department treatment. National estimates are presented by age, sex, and industry divisions. The causes and types of injuries are also described.

Methods

The injury data reported here were extracted from NEISS, which is maintained by the Consumer Product Safety Commission (CPSC). NEISS contains information on product-related injuries collected from a national sample of 91 hospital emergency departments in the United States. Through a collaborative effort between the National Institute for Occupational Safety and Health (NIOSH) and CPSC, data on work-related injuries to persons aged 14 through 17 years have been collected since July 1992. This paper presents data for the first 6 months of surveillance, July through December 1992.

Hospitals participating in NEISS were selected from a probability sample of all hospitals in the United States and its territories, stratified by geographic area and hospital size.[16,17] Variances and confidence intervals were calculated on the basis of weighted estimates,[18] but CPSC recommends that confidence intervals not be calculated for weighted estimates of less than 1200. The numbers presented in the tables represent estimates for the number of injured adolescents in the United States for the entire 6-month period.

A work-related injury was defined as any injury sustained during work performed for pay or other compensation, regardless of eligibility for workers' compensation. Also included were injuries sustained during volunteer work for organized groups such as volunteer fire fighters, hospitals, and charity groups. Persons in the military were excluded from this analysis because employment data comparable to those of civilians were not available and the sample included cases from only one of the five military branches.

Industry information was coded according to the *Standard Industrial Classification Manual.*[19] The nature of the injury was derived directly from the physician's diagnosis.[20]

Although data were not available for the first 6 months of 1992, an estimate for the entire year was extrapo-

The authors are with the Division of Safety Research, National Institute for Occupational Safety and Health, Morgantown, WVa.

Requests for reprints should be sent to Larry A. Layne, MA, Division of Safety Research, National Institute for Occupational Safety and Health, 944 Chestnut Ridge Rd, Morgantown, WV 26505.

This paper was accepted November 15, 1993.

Editor's Note. See related editorial by Pless (p 537) and articles by Castillo et al. (p 646) and Parker et al. (p 606) in this issue.

TABLE 1—Estimated Number and Rates[a] of Adolescent Occupational Injuries in the United States, by Age and Sex, July through December 1992

	Males		Females		Total	
Age, y	No.	Rate	No.	Rate	No.	Rate
14[b]	1028	. . .	274	. . .	1302	. . .
15	2904	6.7	1136	2.6	4040	4.7
16	7084	6.1	4352	4.7	11 436	5.5
17	13 583	7.6	7044	4.7	20 627	6.3
All	24 599	7.0	12 806	4.4	37 405	5.8

Note. Because each number represents one injured adolescent, the numbers were rounded to the nearest integer. Owing to this rounding, the numbers in the rows and columns may not add up to the exact number presented in the totals.
[a]Injury rates were calculated per 100 full-time employees; adolescents injured doing volunteer work were excluded from the rate calculations.
[b]Rates were not calculated because hourly employment data were not available for 14-year-olds.

TABLE 2—Estimated Number of Adolescent Occupational Injuries in the United States, by Body Part Injured and Nature of Injury, July through December 1992

	Nature of Injury						
Body Part	Laceration	Contusion or Abrasion	Sprain or Strain	Burn	Fracture or Dislocation	All Other	Total
Hand or finger	9560	2233	961	1620	614	1592	16 579
Leg, knee or ankle	1189	928	1638	420	356	423	4954
Head, face, or neck	1111	941	34	1092	0	1466	4644
Wrist or arm	853	961	1199	843	342	252	4449
Trunk or shoulder	51	1097	2073	267	77	183	3749
Foot or toe	132	590	137	327	166	818	2171
All other	0	69	0	60	0	732	860
Total	12 896	6819	6042	4629	1555	5465	37 405

Note. Because each number represents one injured adolescent, the numbers were rounded to the nearest integer. Owing to this rounding, the numbers in the rows and columns may not add up to the exact number presented in the totals.

lated from the estimated number of adolescents injured from July through December 1992. To account for the seasonal variability of adolescent work, we assumed that the adolescent injury pattern for 1992 followed the same monthly patterns as for the years 1982 through 1986. Using the monthly distribution of adolescent injuries from previous surveillance, we computed an annual estimate for the entire year of 1992. (NEISS occupational injury data were collected by NIOSH for the years 1982 through 1986. These data were analyzed by month for workers aged 14 through 17 years for the 5-year period. The monthly weighted distribution was used to estimate the number of injuries that occurred during the first 6 months of 1992, a period for which data were not collected.)

In computing injury rates, labor force data from the Bureau of Labor Statistics' Current Population Survey were used as the denominator.[21] The Current Population Survey is a monthly national survey that includes employment information for persons as young as 15 years old. Employment data were converted to full-time employee equivalents.

Injury rates were not calculated for 14-year-olds because of the lack of hourly employment data. Because employment data from the Bureau of Labor Statistics do not include volunteer work, adolescents injured while performing volunteer work were not included in the rate calculations.

This study focuses on nonfatal injuries only. Data from emergency departments tend to underestimate fatal injuries because many victims of traumatic injury die before ever reaching an emergency department.[22] The NEISS system did not include any fatalities to adolescents during the period of this study.

Results

There were 679 occupational injuries to 14- through 17-year-old civilian workers reported in NEISS (July through December 1992). This number represents an estimated 37 405 (95% confidence interval [CI] = 28 000, 46 800) injuries nationwide to adolescent workers for the 6-month period. An estimated 297 (0.8%) of these injured adolescents were performing volunteer work; approximately 948 of the injured youth required hospitalization. It is estimated that about 64 100 (95% CI = 48 100, 80 200) work-related injuries to adolescents were treated in emergency departments during the entire year of 1992.

Table 1 shows the distribution of injuries and injury rates by age and sex for the 6-month period. Males accounted for a larger percentage of injuries (65.8%) and had a higher injury rate than females. The injury rate per 100 full-time employee equivalents for 16- and 17-year-olds combined (for comparison with rates found in other studies) was 7.0 for males and 4.7 for females.

Lacerations were the most common type of injury incurred (Table 2). Contusions and abrasions, sprains and strains, and burns constituted a larger proportion of injuries among females (56.5%) than among males (41.7%), although the injury rates for these three types of injuries were similar for both sexes. However, males' injury rate for fractures and dislocations was twice that of females (0.2 and 0.1, respectively, per 100 full-time employee equivalents); males' laceration rate was nearly twice that of females (2.0 and 1.2, respectively, per 100 full-time employee equivalents).

The hand or finger was the most frequently injured body area (44.3%) (Table 2). For males, 44.5% of injuries were sustained in the hand or finger; the proportion for females was 40.2%. The leg, knee, or ankle accounted for 10.6% of the injuries to males and 18.4% of those to females.

The industry division of retail trade accounted for 53.7% of all injuries and had the highest occupational injury rate

among adolescent workers (Table 3). The largest proportion of injuries in the retail trade industry (71.2%) occurred in eating and drinking establishments (Table 4).

The retail trade industry was characterized by high proportions of lacerations (39.0%) and burns (17.7%) (Table 4). Of laceration injuries in retail trade, a majority resulted from being struck by objects and striking against objects. Hand or finger injuries constituted more than one half of all injuries in retail trade; 84.1% of the injuries to the hand or finger were lacerations. Retail trades accounted for 3553 burn injuries, which primarily resulted from contact with hot objects (73.9%) and exposure to caustic substances (14.8%).

The second-largest number of injuries to adolescents occurred in the service industry. Nearly two thirds of the injuries in the service industry occurred in health services, amusement and recreation, and educational services. Contusion and abrasion injuries were overrepresented in services (30.0%, compared with 15.2% in all other industries) (Table 4).

The agriculture industry accounted for 6.7% of all injuries among adolescents. While lacerations and contusions or abrasions were the most common types of injuries in agriculture, fractures and dislocations were more prevalent in agriculture (14.4%) than in other industries (3.4%) (Table 4). One fifth of all agricultural injuries resulted from being caught in running machinery or equipment.

The manufacturing industry had the second highest injury rate (Table 3) and accounted for 4.0% of all occupational injuries among adolescents. Sprains and strains were the most frequent type of injury in manufacturing (32.3%, compared with 15.5% for all other industries). Forty-seven percent of the sprains and strains in manufacturing were to the hand or finger and 24.8% were to the leg, knee, or ankle.

An examination of the industry divisions by sex shows that a large percentage of injuries incurred by females were in the retail and service industries (Table 3). In agriculture, the proportion of injuries incurred by males was 87.7%, compared with 64.2% for all other industries; also, 40.0% of the injuries in agriculture were to 14- and 15-year-old males, compared with 12.4% for all other industries.

TABLE 3—Estimated Number and Rate[a] of Adolescent Occupational Injuries in the United States, by Industry Division and Sex, July through December 1992

Industry Division	No. Injuries			Injury Rate
	Male	Female	Total	
Retail trade	12 634	7453	20 087	6.3
Services	4049	3632	7681	4.1
Agriculture	2195	308	2502	4.3
Manufacturing	1013	485	1498	5.1
Construction	832	0	832	4.8
Transportation, communication, public utilities	506	34	540	4.4
Wholesale	431	0	431	4.5
Public administration	325	34	359	4.3
Mining[b]	34	0	34	. . .
Finance, insurance, real estate[b]	0	0	0	. . .
Unknown	2580	861	3441	NA
All	24 599	12 806	37 405	5.8

Note. NA = not applicable. Because each number represents one injured adolescent, the numbers were rounded to the nearest integer. Owing to this rounding, the numbers in the rows and columns may not add up to the exact number presented in the totals.
[a]Injury rates were calculated per 100 full-time employees; 14-year-olds and adolescents injured doing volunteer work were excluded from the rate calculations.
[b]Rate was not calculated because of small numbers.

TABLE 4—Estimated Number of Adolescent Occupational Injuries in the United States, by Industry Division and Nature of Injury, July through December 1992

Industry Division	Nature of Injury						
	Laceration	Contusion or Abrasion	Sprain or Strain	Burn	Fracture or Dislocation	All Other	Total
Retail trade	7836	2723	3193	3553	772	2011	20 087
Eating and drinking establishments	5524	1700	2188	3295	416	1177	14 299
Food stores	1270	608	433	132	184	304	2931
General merchandise	524	235	220	0	112	351	1442
All other retail	519	179	352	126	60	180	1416
Services	1772	2304	1258	698	218	1431	7681
Health services	932	807	375	278	0	107	2499
Amusement and recreation	146	308	420	0	150	214	1237
Educational services	340	290	0	270	0	215	1116
Social services	34	244	336	115	0	34	763
Hotels and lodging	60	0	77	0	34	325	495
All other services	261	655	51	34	34	536	1572
Agriculture	890	374	197	17	360	665	2502
Manufacturing	312	150	484	112	43	397	1498
All other	2086	1268	909	249	163	961	5637
Total	12 896	6819	6042	4629	1555	5465	37 405

Note. Because each number represents one injured adolescent, the numbers were rounded to the nearest integer. Owing to this rounding, the numbers in the rows and columns may not add up to the exact number presented in the totals.

Discussion

We estimate that 64 100 of the nation's adolescents were treated in emergency departments for occupational injuries in 1992, and that adolescents have an injury rate of 5.8 per 100 full-time employee equivalents. These national estimates must be considered conservative because they are based on a

study of only injuries treated in emergency departments. Previous studies have found that about 36% of all work-related injuries are treated in emergency departments,[15] and fatal injuries are rarely included in emergency department surveillance.[22]

Adolescent occupational injuries can be prevented only once hazards have been identified and age-specific intervention strategies have been developed and incorporated into the industry's safety and training programs. Injuries reported in retail eating and drinking establishments accounted for the largest number of injuries in any industry subcategory, a finding supported by a study of workers' compensation cases in the state of New York.[9] Many of these injuries, particularly burns and lacerations, occurred in fast food establishments. Implementing methods to prevent burns in eating and drinking establishments, such as those previously recommended on the basis of detailed work-site inspections,[23,24] could prevent about 5650 adolescent occupational burns each year. Detailed studies and work-site investigations could identify prevention strategies in other sectors of retail trade,[11] services, manufacturing, and agriculture.

Although the injury rate in agriculture was not high in comparison with those in other industries, there was some indication that these injuries were more severe. Fractures and dislocations accounted for a higher percentage of the injuries in agriculture than in other industries. This finding is similar to findings from studies of adolescent workers' compensation claims.[9,12] Agricultural occupations declared hazardous by the secretary of labor are sanctioned for youth as young as 16 (and for even younger children working on their parents' farms).[25] In contrast, the minimum age for work declared hazardous in nonagricultural occupations is 18.[26]

There are many workplace risk factors that are unique to the adolescent workforce. Hayes-Lundy et al. found that length of employment is not related to injury experience among youth.[23] In addition, adolescents may be less familiar than adults with workplace hazards and safety awareness. Working youth, their parents, and employers must be informed about child labor laws, the risk of injury to adolescent workers, and the workplace hazards that exist in industries where adolescents are employed. □

Acknowledgments

The authors gratefully acknowledge John Myers and Dr Terry Wassell for their statistical consultation; Dr Philip Landrigan and Dr David Parker for their reviews of this paper; and Art McDonald and Eileen Kessler of CPSC for their support in data collection and methodological review.

References

1. Richter ED, Jacobs J. Work injuries and exposures in children and young adults: review and recommendations for action. *Am J Ind Med.* 1991;19:747–769.
2. Greenberger E, Steinberg L. *When Teenagers Work: The Psychological and Social Costs of Adolescent Employment.* New York, NY: Basic Books In ; 1986.
3. Stephens W. *Our Children Should Be Working.* Springfield, Ill: Charles C Thomas; 1979.
4. Steinberg L, Dornbusch SM. Negative correlates of part-time employment during adolescence: replication and elaboration. *Dev Psychol.* 1991;27:304–313.
5. Castillo DN, Landen DD, Layne LA. Occupational injury deaths of 16- and 17-year-olds in the United States. *Am J Public Health.* 1994;84:646–649.
6. Suruda A, Halperin W. Work-related deaths in children. *Am J Ind Med.* 1991;19:739–745.
7. Brooks DR, Davis LK, Gallagher SS. Work-related injuries among Massachusetts children: a study based on emergency department data. *Am J Ind Med.* 1993;24:313–324.
8. Glor ED. Survey of comprehensive accident and injury experience of high school students in Saskatchewan. *Can J Public Health.* 1989;80:435–440.
9. Belville R, Pollack SH, Godbold JH, Landrigan PJ. Occupational injuries among working adolescents in New York State. *JAMA.* 1993;269:2754–2759.
10. Parker DL, Carl W, French LR, Martin FB. Characteristics of adolescent work injury reported to the Minnesota Department of Labor and Industry. *Am J Public Health.* 1994;84:606–611.
11. Banco L, Lapidus G, Braddock M. Work-related injury among Connecticut minors. *Pediatrics.* 1992;89:957–960.
12. Heyer NJ, Franklin G, Rivara FP, Parker P, Haug JA. Occupational injuries among minors doing farm work in Washington State: 1986 to 1989. *Am J Public Health.* 1992;82:557–560.
13. Parker DL, Clay RL, Mandel JH, Gunderson P, Salkowicz L. Adolescent occupational injuries in Minnesota: a descriptive study. *Minn Med.* 1991;74:25–28.
14. Schober SE, Handke JL, Halperin WE, Moll MB, Thun MJ. Work-related injuries in minors. *Am J Ind Med.* 1988;14:585–595.
15. Coleman PJ, Sanderson LM. Surveillance of occupational injuries treated in hospital emergency rooms—United States, 1982. *MMWR.* 1983;32:31SS–37SS.
16. *The NEISS Sample (Design and Implementation).* Bethesda, Md: US Consumer Product Safety Commission, Division of Hazard and Injury Data Systems; March 1986 (revised October 1992).
17. Marker D, Waksberg J, Branden J. *NEISS Sample Update: Final Report.* Rockville, Md: Westat, Inc.; June 3, 1988. Publication CPSC-C-87-1144.
18. *National Electronic Injury Surveillance System (NEISS): Estimated Generalized Relative Sampling Errors.* Bethesda, Md: US Consumer Product Safety Commission; September 1988.
19. *Standard Industrial Classification Manual, 1987.* Washington, DC: Office of Management and Budget; 1987.
20. *National Electronic Injury Surveillance System Coding Manual.* Bethesda, Md: US Consumer Product Safety Commission; 1985.
21. *Annual Average Supplements: Employment and Earnings.* Washington, DC: US Bureau of Labor Statistics; 1993:40(issue no. 1).
22. Rivara FP. Fatal and nonfatal farm injuries to children and adolescents in the United States. *Pediatrics.* 1985;76:567–573.
23. Hayes-Lundy C, Ward RS, Saffle JR, Reddy R, Warden GD, Schnebly WA. Grease burns at fast-food restaurants: adolescents at risk. *J Burn Care Rehabil.* 1991;12:203–208.
24. Centers for Disease Control and Prevention. Occupational burns among restaurant workers—Colorado and Minnesota. *MMWR.* 1993;42:713–716.
25. *Child Labor Requirements in Agriculture under the Fair Labor Standards Act* (Child Labor Bulletin no. 102). Washington, DC: US Dept of Labor, Employment Standards Administration, Wage and Hour Division; 1984. WH publication 1295.
26. *Child Labor Requirements in Nonagricultural Occupations under the Fair Labor Standards Act* (Child Labor Bulletin no. 101). Washington, DC: US Dept of Labor, Employment Standards Administration, Wage and Hour Division; 1978. WH publication 1330.

Risk Factors for Violent Behavior in Elementary School Boys: Have You Hugged Your Child Today?

Jonathan L. Sheline, MD, MS, Betty J. Skipper, PhD, and W. Eugene Broadhead, MD, PhD

ABSTRACT

Seventeen Hispanic elementary schoolboys with violent behavior problems were compared with 27 matched control students who were not overtly violent at school. Violent boys were significantly more likely to not live with their fathers, to have unmarried parents, to have more siblings, and to have fathers who never show them affection. Parents of violent boys were more likely than those of matched control students to use spanking for discipline and to admit that they rarely express affection for their sons. (*Am J Public Health.* 1994;84:661–663)

Introduction

Violence is an especially serious public health problem in the United States, as evidenced by our homicide rate of 8.9 per 100 000 in 1988, the highest in the industrialized world (i.e., all of Europe, North America, southern South America, and the Pacific Rim countries).[1] Elementary school is one of the earliest settings in which we can study violent misbehavior. Longitudinal studies show that school truancy, fighting, or troublesome behavior as early as the first grade predicts similar behavior in high school[2] and predicts arrests for juvenile delinquency.[3–5] Juvenile delinquency, in turn, is associated with higher adult rates of arrests, violence, and drug use.[6–12]

Longitudinal studies of criminals find that the parenting practices associated with future delinquency include poor supervision or neglect by either parent,[13–17] overly harsh discipline by either parent,[10,13,14,16] and lack of affection by the mother[14,15] and/or father.[16] Despite this work, however, "the causal connection between the family and delinquency remains ambiguous and equivocal."[18(p263)] Most studies to date have suffered from the lack of a specific comparison group, the use of composite measures of family dysfunction instead of specific parenting practices,[11,14,15,19] and the use of legally defined juvenile delinquency as the outcome of interest when preventive intervention may be too late.

We report on a study of young schoolchildren, whose violent behavior was identified by their teachers rather than the court system, and of a rigorously matched control group. We studied minority, inner-city children, a population that has been neglected by all but one recent study.[20,21] Our purpose was to determine what specific parenting practices and family characteristics are associated with violent behavior in young schoolchildren.

Methods

Study Population

The study was conducted during the 1988/89 school year at a 530-student public elementary school in an inner-city, low-income, Hispanic neighborhood in Albuquerque, NM. A case–control study design was used. Case subjects were defined as students who received two or more infraction notices from teachers during the first semester of the school year for any of seven violent misbehaviors, including fighting, bringing dangerous items to school, and using gangs to intimidate others.

Two control students per case subject were selected from class lists, using the alphabetically nearest names that fit the matching criteria of having the same teacher, sex, ethnicity, and age within 1 year. Potential control students were excluded if they, or any of their siblings at the school, had been written up for a violent offense during the fall of 1988, or if they had enrolled late in school and had less time to accumulate offenses.

Questionnaires

All students in grades two through five filled out a 65-item questionnaire during class time, administered by their teachers with standard written instructions. Most questions used a Likert scale with three choices. Twenty-eight of the 65 items were adapted from a previously validated adolescent health survey[22]; the remainder were developed by the investigators. The questionnaire was field tested and revised using a convenience

Jonathan L. Sheline and W. Eugene Broadhead are with the Department of Community and Family Medicine, Duke University Medical Center, Durham, NC. Betty J. Skipper is with the Department of Family and Community Medicine, University of New Mexico School of Medicine, Albuquerque.

Requests for reprints should be sent to Jonathan L. Sheline, MD, MS, Box 2914, Durham, NC 27710.

This paper was accepted November 12, 1993.

TABLE 1—Student Questionnaire Variables with Significant[a] Difference in Responses of Case vs Control Subjects

Compared with Matched Control Students, Children Who Exhibit Violent Misbehavior in School Are More Likely to	Proportion[a] Case Subjects (n = 17)	Control Subjects (n = 27)	Matched Odds Ratio[b]	P
Not live with fathers	.53	.11	10.9	.026
Have parents who are not married	.59	.15	6.0	.022
Have fathers or male guardians who never show them physical affection	.31[c]	.00[d]	. . .[e]	.017[f]
Have fathers or male guardians who never express pride in or affection for them	.31[c]	.00[d]	. . .[e]	.017[f]
Report that they got at least one "bad note"	1.00	.56	. . .[e]	.0016[f]
Have more siblings (average no.)	4.1	2.4	NA	.026[g]

[a]Denominators may vary slightly; students were deleted from the study if they did not answer the question or if they had no match for the matched analysis.
[b]Calculated by conditional logistic regression.
[c]n = 16.
[d]n = 24.
[e]Matched analysis was not possible since no control is "exposed" or else all cases are exposed (in which case, the odds ratio would be infinity).
[f]Fisher's Exact Test, two sided.
[g]Student's *t* test, two sided.
*$P < .05$, two sided.

TABLE 2—Parenting Practices of Primary Caretakers of Violent Children vs Those of Matched Control Students

Behavior of Caretakers	Proportion[a] Case Subjects (n = 15)	Control Subjects (n = 20)	Matched Odds Ratio[b]	95% Confidence Interval
Occasionally or never express pride or affection to child	0.73	0.25	10.0	1.23, 86.0
Use spanking for discipline when child misbehaves at home	0.73	0.35	8.5	1.01, 71.0
Are not able to identify three things they like about child	0.53	0.21[c]	7.6	0.92, 64.0
Have no help taking care of children	0.27	0.10	4.4	0.44, 43.0
Occasionally or never show child physical affection	0.40	0.10	4.0	0.79, 20.0
Are not interested in discussing parenting skills with school counselor	0.57[d]	0.37[c]	3.2	0.62, 16.3
Are interested in discussing child's behavior with school counselor	0.71[d]	0.50[e]	2.4	0.44, 12.6
Let child watch more than 3 hours of television on school nights	0.40	0.25	1.9	0.43, 8.0

[a]Denominators vary slightly; caretakers were deleted from the study if they did not answer the question or if they had no match for the matched analysis.
[b]Calculated by conditional logistic regression.
[c]n = 19.
[d]n = 14.
[e]n = 18.

sample of two students from each grade. Kindergarten and first-grade students could not reliably answer the questionnaire and were excluded from the study.

In-home interviews using a 21-item questionnaire were sought from a parent or guardian of every case and control subject, and were conducted by six trained schoolteachers. The interviewers were blind to the case or control status of the child in 39 of 41 interviews.

Both questionnaires were developed in consultation with an adolescent psychiatrist and are available on request.

Analysis

Matched odds ratios were calculated with conditional logistic regression. Matching was broken for comparisons of continuous variables. When no control subjects were exposed to a given risk factor, Fisher's Exact Test was used.

Results

Study Participants

During the 78 school days of the first semester, 74 boys and 5 girls were given infraction notices for violent behavior. Of those, 48 boys and 4 girls got only one notice for a violent offense and were thus excluded from the study to provide a clearer demarcation between case and control subjects. After 4 children in kindergarten and first grade were excluded, 22 boys and 1 girl remained eligible for study. Six of the 23 had moved, including the 1 girl, leaving 17 male case subjects.

Two control students per case subject were desired, but six case subjects had only one classmate who met the matching and exclusion criteria for controls, and one control student was excluded later because of an incomplete questionnaire; this left 27 control students. Parent interviews were completed by the parent or guardian of 16 of the 17 case subjects and 25 of the 27 control students.

Risk Factors for Violent Behavior

Compared with matched control students, children who exhibited violent misbehavior in school were 11 times as likely not to live with their fathers and 6 times as likely to have parents who were not married; they also had almost twice as many siblings. Their fathers (or male guardians, for those who did not live with their fathers) were significantly more likely never to show them physical affection and never to express pride in or affection for them (Table 1).

Responses to questions on parenting practices were counted only if the parent or guardian interviewed said he or she was the primary caretaker of the child, thereby yielding 15 case and 20 control subjects. Table 2 shows re-

sponses for all questions on parenting practices. The primary caretakers of the 15 violent children were significantly more likely than those of the 20 matched control students to report that they rarely or never expressed pride or affection to their child and that they used spanking for discipline when the child misbehaved at home.

Discussion

Early violent behavior in the second through fifth grades is correlated with both structural and functional family characteristics in this low-income, largely Hispanic population. The small sample size limits this study's power to detect significant differences. The cross-sectional ascertainment of exposures precludes inferences about cause and effect.

The finding that violent boys got less affection from their male guardians may be confounded by the finding that violent boys were much more likely not to be living with their fathers. Indeed, it has been shown that stepparenthood is the single most powerful risk factor for child abuse.[23] When the analysis was restricted to only those boys who lived with their biologic fathers, violent boys were still almost twice as likely as matched control subjects to have fathers who never hugged them or expressed verbal affection, but this was not statistically significant.

This study focuses on a minority population and addresses aspects of family influences on high-risk behaviors at a very early age when intervention may be feasible. Our data show that boys from families with absent fathers, divorced parents, or numerous siblings are at higher risk for violent behavior. Recent studies have shown that family demographic factors, including single parents and low socioeconomic status,[24,25] have little or no effect on future delinquency after controlling for poor parenting.[16,26,27] The parenting practices most strongly associated with violent behavior in our study were lack of affection shown by either parent, especially fathers, and use of spanking for discipline. Intervention trials are urgently needed to evaluate how programs that encourage and enable parents to show more affection affect children's behavior and to teach parents to use disciplinary measures other than spanking. □

Acknowledgments

This research was supported by a Health of the Public grant to the University of New Mexico from the Pew Charitable Trusts and the Rockefeller Foundation, and by a US Public Health Service grant, 5 D32 PE14010–12, to Duke University.

The authors would like to express their warmest appreciation to the students and staff of the elementary school for their enthusiastic participation in this research. Special thanks go to the school's nurse, Mrs Gloria Sandoval, RN, and to administrative assistant Mrs Carol Vigil. Thanks also go to Drs Arthur Kaufman, Lloyd Michener, Maurice A. Hitchcock, and Ross Snyder for their suggestions and review of the manuscript.

References

1. *1991 Demographic Yearbook.* New York, NY: United Nations; 1992. (Crude death rates reported; years of report: 1987 to 1990.)
2. Robins LN, Ratcliff KS. The long-term outcome of truancy. In: Hersov L, Berg I, eds. *Out of School: A Modern Perspective in Truancy and School Refusal.* New York, NY: John Wiley & Sons; 1980.
3. Mulligan DG, Douglas JWB, Hammond WA, Tizard F. Delinquency and symptoms of maladjustment. *Proc Royal Soc Med.* 1963;56:1083–1086.
4. Craig MM, Glick SJ. School behavior related to later delinquency and nondelinquency. *Criminologica.* 1968;5:17–27.
5. Farrington DP, West DJ. A comparison between early delinquents and young aggressives. *Br J Criminol.* 1971;11:341–358.
6. Kandel DB, Ravies VH, Kandel PI. Continuity in discontinuities: adjustment in young aduithood of former school absentees. *Youth Soc.* 1984;15:325.
7. Robins LN. *Deviant Children Grown Up: A Sociological and Psychiatric Study of Sociopathic Personality.* Huntington, NY: Robert E. Krieger Publishing Co; 1974.
8. Robins LN. Sturdy childhood predictors of adult antisocial behaviour: replications from longitudinal studies. *Psychol Med.* 1978;8:611–622.
9. Osborn SG, West DJ. The effectiveness of various predictors of criminal careers. *J Adolesc.* 1978;1:101–117.
10. Farrington DP. Environmental stress, delinquent behavior, and convictions. In: Sarason IG, Spielberger CD, eds. *Stress and Anxiety.* Washington, DC: Hemisphere; 1979;6:chap 5.
11. Wadsworth M. *Roots of Delinquency: Infancy, Adolescence, and Crime.* Oxford, England: Martin Robertson & Co; 1979.
12. Gersten JC, Langner TS, Eisenberg JS, Simcha-Fagan D, McCarthy ED. Stability and change in types of behavioral disturbances of children and adolescents. *J Abnorm Child Psychol.* 1976;4:111–127.
13. Farrington DP. The family backgrounds of aggressive youths. In: Hersov LA, Berger M, Shaffer D, eds. *Aggression and Antisocial Behavior in Childhood and Adolescence.* Oxford, England: Pergamon Press; 1978:73–93.
14. McCord J. Some child-rearing antecedents of criminal behavior in adult men. *J Pers Soc Psychol.* 1979;37:1477–1486.
15. Craig MM, Glick SJ. Ten years' experience with the Glueck social prediction table. *Crime and Delinquency.* 1963;9:249–261.
16. Laub JH, Sampson RJ. Unraveling families and delinquency: a reanalysis of the Gluecks' data. *Criminology.* 1988;26:355–380.
17. Widom CS. The cycle of violence. *Science.* 1989;244:160–166.
18. Wells LE, Rankin JH. Direct parental controls and delinquency. *Criminology.* 1988;26:263–285.
19. Rutter M. Family, area, and school influences in the genesis of conduct disorders. In: Hersov LA, Berger M, Shaffer D, eds. *Aggression and Antisocial Behavior in Childhood and Adolescence.* Oxford, England: Pergamon Press; 1978: 95–113.
20. Wishner AR, Schwarz DF, Grisso JA, Holmes JH, Sutton RL. Interpersonal violence-related injuries in an African-American community in Philadelphia. *Am J Public Health.* 1991;81:1474–1476.
21. Grisso JA, Wishner AR, Schwarz DF, Weene BA, Holmes JH, Sutton RL. A population-based study of injuries in inner-city women. *Am J Epidemiol.* 1991; 134:59–68.
22. Blum R, Harris LJ, Resnick MD, Rosenwinkel K. Technical report on the Adolescent Health Survey. Minneapolis, Minn: University of Minnesota Hospital and Clinic, Adolescent Health Program; June 1989.
23. Daly M, Wilson M. *Homicide.* Hawthorne, NY: Aldine de Gruyter; 1988.
24. West DJ. *Present Conduct and Future Delinquency.* New York, NY: International Universities Press, Inc; 1969.
25. Wolfgang ME, Figlio RM, Sellin T. *Delinquency in a Birth Cohort.* Chicago, Ill: University of Chicago Press; 1972.
26. Larzelere RE, Patterson GR. Parental management: mediator of the effect of socioeconomic status on early delinquency. *Criminology.* 1990;28:301–323.
27. Cernkovich SA, Giordano PC. Family relationships and delinquency. *Criminology.* 1987;25:295–321.

• The central factor in increased risk was the high frequency of sexual relationships outside the regular partner relationship. These "steppin' out" relationships contributed to considerable "gamesmanship" in both women and men, covering up both the behavior and resulting sexually transmitted diseases. Yet the quantitative data indicated that people were far more optimistic about an intervention effort aimed at reducing sexual partners than they were about increasing condom use (although this result was stronger among women).

On the basis of these results, a community-based, peer-oriented program for small, mixed-gender groups was designed. The aim of the program is to reduce the number of sexual partners and to strengthen the primary relationship when one is present. Trained peer counselors will facilitate small group discussions, presentations, and activities aimed at increasing individual self-esteem, individual and mutual responsibility, and understanding between men and women. The intervention program will provide information about sex and sexual health and present positive role models and relationship models for men and women. Approximately $100 000 in funding is being sought for a 2-year period to develop, test, evaluate, and modify three rounds of the intervention.

The delineation of focal subareas and subpopulations permits a broader range of methodological tools that can be used by local health department staff and can be linked more directly to program development. These tools include the collection of qualitative data directly from members of the target population, linked to the use of focused surveys with a relatively small sample size. The combination of qualitative and quantitative data provides mutual verification, stimulates new lines of discovery, and improves the specificity of data collection methods and instruments. It requires less time, personnel, and expense than more traditional methods of community health evaluation and need assessment. The development and testing of the method were carried out as part of a larger contract between the University of Connecticut and the Connecticut Department of Health Services and accounted for $40 000 of the larger contract. Personnel included a half-time project coordinator, four part-time key informant interviewers who were members of the targeted subpopulation, and three student data processors who abstracted information from the sexually transmitted disease reports.

The methodology encourages the participation of many sectors of the community, including the funding source, the sponsoring organization, other community-based service providers, and the focal population. Maximizing the participation of each of these sectors in the research process contributes to their commitment to the results and to the development and implementation of an intervention. The data collection procedures called for require input from a trained researcher (we would suggest the involvement of local university faculty) but maximize the involvement of health department and community people in design, data collection, and analysis.

This research approach can serve to redefine the role of the university as a community partner and the role of the local health department not only as an effective user of data but as a generator of data in its search for the local solutions that are so much a part of public health tradition. □

Stephen L. Schensul, PhD
Merrill R. Eisenberg, PhD
John Glasgow, PhD
John Huettner, MS

Requests for reprints should be sent to Stephen L. Schensul, PhD, Department of Community Medicine, MC-1910, University of Connecticut Health Center, Farmington, CT 06030.

The Rhode Island Sun Smart Project: Skin Cancer Prevention Reaches the Beaches

Skin cancer is by far the most common malignancy in the United States. More than 600 000 Americans will be affected this year, exceeding the number affected by the next four most common sites combined.[1] Solar ultraviolet radiation has been identified as the major cause of the epidemic. Because most ultraviolet radiation exposure is intentional, primary efforts to prevent skin cancer must focus on encouraging individual precautionary behaviors. The beach in summer is a site of special concern, inasmuch as exposure to solar ultraviolet radiation is particularly intense for many people at this time and place.

The purpose of the Rhode Island Sun Smart Project was to deliver skin cancer prevention and health promotion efforts directly to sunbathers at Rhode Island (population: 1 000 000) beaches. A 1-year grant of $35 000 from the National Cancer Institute provided initial funding for the development of measurement models and questionnaires based on the transtheoretical model of behavior change.[2-4] A subsequent 2-year grant of $50 000 from the Rhode Island Cancer Prevention Research Consortium provided funding for the development of interventions and feasibility testing of the beach as an intervention site. Interventions were conducted at one of three Rhode Island state beaches on 11 days during the summers of 1991 and 1992. Rain, forecasts of rain (resulting in poor beach attendance), cold weather, and a hurricane caused us to abandon attempts on 9 additional days in 1991 and 2 days in 1992. Project staff included a health psychologist, a dermatologist, a graduate research assistant, a project coordinator, and six trained interviewers.

Seven interventions were developed and tested:

1. An *educational pamphlet* (*Skin Care under the Sun*,[5] developed by the American Academy of Dermatology) described the hazards of sun exposure and appropriate precautionary behaviors. Pamphlets were distributed in conjunction with all other interventions.

2. Free *sunscreen samples* were distributed to beachgoers with verbal instructions from the interviewers on proper use. The samples were donated to the American Cancer Society by the manufacturer and consisted of 4-oz bottles of a waterproof sunscreen in an emollient base with a sun protection factor (SPF) rating of 15.

3. The *sun sensitivity assessment and feedback* intervention consisted of a few simple questions that assessed intrinsic ("phenotypic") sun sensitivity, an important indicator of skin cancer risk.[6] Feedback describing appropriate preventive actions specific to the participant's degree of risk (high, moderate, or low) was preprinted on adhesive labels and attached to the educational pamphlet.

4. The *sun scanner* intervention consisted of a mirror and a "black light" (Wood's light) in a cloth-enclosed box into which the participant looked. Wood's light is a low-intensity source of blue and near-ultraviolet light that accen-

tuates epidermal pigmentation, dramatizing the extent of ultraviolet damage and photoaging—the cosmetic consequences of chronic sun exposure on the face. The scanner was loaned to the Brown University Dermatoepidemiology Unit free of charge from the manufacturer through the American Academy of Dermatology.

5. Both *educational* and *dramatic videotape* interventions were used. *Skin Care under the Sun* (American Academy of Dermatology) is a 7-minute staged discussion between a mother and her teenaged daughter about the dangers of sun exposure and the necessity for precautions. *Goodbye Sunshine,* a 20-minute segment from the Australian *60 Minutes* television program, recounts the story of a young man who devoted himself to informing the public about the dangers of sun exposure and the importance of early detection of skin cancer until his death from melanoma.

6. The *skin microtopography* intervention required making a plastic impression, similar to a dental impression, of the skin in the crow's-feet area near the eye. These impressions can be graded according to the degree of wrinkling, a manifestation of sun-induced aging of the skin. We chose not to implement this intervention after pilot testing indicated that it would be very difficult to conduct in the beach environment.

7. The *ultraviolet* and *polarized light instant photography* interventions were developed for year 2 of the project to circumvent limitations of the sun scanner and microtopography interventions, respectively. The ultraviolet light camera was used to take a full frontal black-and-white photograph of the face, revealing epidermal pigmentation damage similar to that shown by the sun scanner. The polarized light camera was used to take a close-up color photograph of the crow's-feet area that revealed structural damage to the skin.

Each intervention was conducted at least twice at different beaches and on different days. The selected beaches had distinct "personalities": a teen beach, a family beach, and a community beach. Permission for beach access was obtained from the Rhode Island Department of Environmental Management. Beach staff at all sites were very supportive, often asking to participate in the interventions and volunteering to reorganize the available space to accommodate our needs.

A $500 lottery incentive was offered to recruit participants, although many beachgoers reported that the lottery did not influence their decision to participate. The interviewers were easily spotted on the beach; they wore matching white hats, sunglasses, bright yellow T-shirts printed with the Sun Smart logo and the names of the project sponsors (the University of Rhode Island, Brown University, and the American Cancer Society). The T-shirts legitimized the project while serving as an additional informative intervention. The Sun Smart logo consists of a sun (located in the center of the T-shirt) that has in each ray a slogan or word associated with safe sun behavior (*sunglasses, Slip! Slop! Slap!, SPF15+, hat, shirt*). The intervention staff modeled appropriate sun behavior, captured attention, and promoted interaction. The local broadcast and print media took notice of the project, resulting in coverage on evening news broadcasts and follow-up newspaper stories.

A total of 1016 beachgoers were contacted during the 11 days of testing. Overall participation was high (62%, n = 633), with a median acceptance rate of 89% (range: 21%–96%). Participation was higher than 80% for all interventions except the videotapes (see Table 1). Although videotape would seem to be an important and convenient means of public health education, there are barriers to its use at the beach. These barriers include the need for access to electrical power in an area out of direct sunlight, the requirement that participants leave the beach and come to the pavilion area for an extended period of time, and the inevitable delays (which discouraged participation) for some individuals because the videotape cannot be shown until a reasonable number of people have gathered to watch. In addition, the only beach well-equipped for this intervention (the teen beach) also had the lowest acceptance rates for other interventions.

Participation was much higher for the remaining interventions. The acceptance rate for the sun scanner is especially noteworthy, as the scanner shared many of the limitations of the videotapes, such as the need for access to electrical power and the requirement that sunbathers leave their place on the beach to go to the pavilion area. The sun scanner drew much curiosity, dramatic reactions, and many questions from participants. The sun sensitivity assessment required only 2 minutes to complete, was delivered directly to participants on the beach, and provided immediate feedback. The sunscreen intervention was also well accepted; it was easy to deliver, it was free to participants, and it gave immediate access to sun protection.

TABLE 1—Participation Rates for Beach-Based Interventions (n = 633)

	No.	Acceptance, %[a]
Intervention		
Photographs	80	95
Pamphlets	98	95
Sun sensitivity assessment	137	90
Sun scanner	76	83
Sunscreen	126	81
Videotapes	116	27
Beaches		
Community	164	92
Family	121	86
Teen	348	50

[a]Percentage of individuals offered the intervention who agreed to participate.

The highest participation rates were obtained for the pamphlet and photography interventions. The photography interventions, especially the ultraviolet light photographs, generated a great deal of interest among sunbathers. Although in some ways not quite as compelling as the "real time" image provided by the sun scanner, the photographs had several important advantages. In particular, the photos were directly deliverable to sunbathers on the beach and provided a convenient and highly personalized take-home reminder of the consequences of unprotected exposure to solar ultraviolet radiation. Furthermore, project participants who showed their photographs to family and friends help to disseminate an important public health message.

Participation rates also varied by beach site (see Table 1). The relatively low rate for the teen beach underscores the difficulty of reaching the adolescent population, an important target group for "safe sun" messages.

On the basis of our experience, we recommend a beach intervention package consisting of pamphlets, sunscreens, sun sensitivity assessments, and photographs. Excluding the research and

evaluation components of our project, we estimate a cost per participant of about $30. Costs could be reduced by soliciting contributions of pamphlets and sunscreens and by enlisting the cooperation of local departments of health or volunteer agencies such as the American Cancer Society. Economy of scale would offset relatively costly one-time expenses (e.g., equipment). We are also exploring the development of computer-based expert systems to provide follow-up interventions for the home.[7] Pending continued funding, we intend to conduct further pilot testing of our intervention package.

The Sun Smart Project has demonstrated the feasibility of the beach as a site for skin cancer prevention and health promotion efforts that reach individuals at high risk for skin cancer. Perhaps our biggest concern in conducting this project was that beachgoers would not want to be bothered while spending a relaxing day in the sun with family and friends. In fact, many sunbathers were bored and looking for something to do. The Sun Smart Project also provided valuable data pertinent to assessing which interventions are viable at the beach and identifying barriers to implementation in larger public health efforts. Results continue to be analyzed to assess readiness of the interventions for public health dissemination. □

Joseph S. Rossi, PhD
Linelle M. Blais, PhD
Martin A. Weinstock, MD, PhD

Joseph S. Rossi and Linelle M. Blais are with the Cancer Prevention Research Center, University of Rhode Island, Kingston. Martin A. Weinstock is with the Dermatoepidemiology Unit, Veterans Affairs and Roger Williams Medical Center and Brown University, Providence, RI.

Requests for reprints should be sent to Joseph S. Rossi, PhD, Cancer Prevention Research Center, Department of Psychology, Flagg Rd, University of Rhode Island, Kingston, RI 02881-0808.

More information about the project may be obtained from Dr Rossi.

Acknowledgments

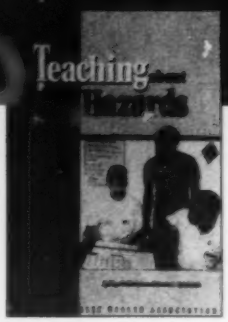
This research was supported by grants CA50087 and CA51690 from the National Cancer Institute and by a developmental grant from the Rhode Island Cancer Prevention Research Consortium.

We thank the following: the American Cancer Society, Rhode Island Division, and Johnson & Johnson for the contribution of sunscreens; the American Academy of Dermatology and Schering-Plough for the loan of the sun scanner; Margaret Boyle for overall project coordination and the design of the Sun Smart logo and T-shirts; and Gabrielle Reed and Sally Cottrill for their help on year 2 of the project.

References

1. Boring CC, Squires TS, Tong T. Cancer statistics, 1991. *Cancer.* 1991;41:19–36.
2. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change: applications to addictive behaviors. *Am Psychol.* 1992;47:1102–1114.
3. Rossi JS. Exploring behavioral approaches to UV risk reduction. In: Moshell A, Blankenbaker LW, eds. *Sunlight, Ultraviolet Radiation, and the Skin.* Bethesda, Md: National Institutes of Health; 1989:91–93.
4. Rossi JS. The hazards of sunlight: a report on the Consensus Development Conference on Sunlight, Ultraviolet Radiation, and the Skin. *Health Psychologist.* 1989;11(3):4–6.
5. *Skin Care under the Sun.* Evanston, Ill: American Academy of Dermatology; 1958. Pamphlet.
6. Weinstock MA. Assessment of sun sensitivity by questionnaire: validity of items and formulation of a prediction rule. *J Clin Epidemiol.* 1992;45:547–552.
7. Velicer WF, Prochaska JO, Bellis JM, et al. An expert system intervention for smoking cessation. *Addict Behav.* 1993;18:269–290.

The Practical Guide to Teaching Workers about Job Hazards



Teaching about Job Hazards: A Guide for Workers and Their Health Providers

By Nina Wallerstein, DrPH, and Harriet Rubenstein, JD, MPH.
170 pages, softcover, 1993

Occupational disease and injury are vast and growing health problems; however, they are almost 100% preventable with appropriate worker education in health and safety. This practical manual gives guidelines on providing health and safety education to workers who face on-the-job hazards. It contains actual worksheets, tips, case studies, and ideas that can be used immediately. Intended for readers who are already familiar with occupational health.

Contents: Guidelines for Worker Education; Educating Individual Worker-Patients; Developing a Training Workshop; Providing Education in Conjunction with Screenings; Preparing Factsheets and Training Materials; Developing Training Sessions for Workers with Low Literacy and Limited English Proficiency; Teaching Adults: Guidelines for Using Various Methods; Evaluating Health and Safety Education; Additional Prevention Activities; Occupational Health and Training Resources; Sample Training Materials: Factsheets, Activities, Training Agendas; Recommendations for a Basic Worker and Community Occupational Health Library; A Basic Occupational Health Library for Primary Care Providers. ISBN 0-87553-209-8

APHA Publications Sales
Department 5037
Washington, DC 20061-5037

Stock No. 008H $20.00 Nonmembers $14.00 APHA Members

Shipping and Handling: U.S. $4.00 per book; Foreign $6.50 per book. All orders must be prepaid with check drawn on U.S. banks. Do not send cash. Call (202) 789-5667 for volume discounts.

Charge by phone: Call (202) 789-5636 to charge to VISA or MASTERCARD.

1/93

A2623-1

American Journal of Public Health

KIANKA
HEALTH REPORTER
CONGRESSIONAL INFORMATION SVC
4520 EAST-WEST HIGHWAY
BETHESDA MD 20814
9375

May 1994, Vol. 84, No. 5 Established 1911

American Public Health Association

Featuring
Diet and Nutrition
Public Health Research

Editorials and Annotations

713 Editorial: The Politics of Dietary Guidance—A New Opportunity *M. Nestle*
715 Editorial: Ecologic Analysis as Outlook and Method *C. Poole*
717 Editorial: The Paper Route for Submissions to the Journal *M. E. Northridge and M. Susser*
718 Annotation: Seven Fatal Flaws in Submitted Manuscripts *M. E. Northridge and M. Susser*
720 Annotation: Insulating Public Health from Extremist Politics—Do We Need Boards of Health? *D. P. Francis*

Commentary

722 *Trans* Fatty Acids: Are the Effects Only Marginal? *W. C. Willett and A. Ascherio*

Public Health Policy Forum

725 Porches, Politics, and Public Health *D. R. Smith*
727 Health Care Reform—More than Cost Containment and Universal Access *E. Feingold*
729 The Future of Public Health in Health Care Reform *V. Navarro*

Articles

VARIOUS TOPICS

731 Chronic Neurological Sequelae to Organophosphate Pesticide Poisoning *K. Steenland, B. Jenkins, R. G. Ames, M. O'Malley, D. Chrislip, and J. Russo*
737 Health Status Differentials in the People's Republic of China *J. S. Lawson and V. Lin*
742 Pap Smear and Mammogram Screening in Mexican-American Women: The Effects of Acculturation *L. Suarez*
747 Risk Behavior and Correlates of Risk for HIV Infection in the Dallas County Household HIV Survey *G. M. McQuillan, T. M. Ezzati-Rice, A. B. Siller, W. Visscher, and P. Hurley*
754 Self-Reports of HIV Risk Factors by Patients at a Sexually Transmitted Disease Clinic: Audio vs Written Questionnaires *B. O. Boekeloo, L. Schiavo, D. L. Rabin, R. T. Conlon, C. S. Jordan, and D. J. Mundt*
761 HIV Vaccine Trials: Will Intravenous Drug Users Enroll? *K. Meyers, D. S. Metzger, H. Navaline, G. E. Woody, and A. T. McLellan*
767 A Community-Based Heart Disease Intervention: Predictors of Change *M. A. Winkleby, J. A. Flora, and H. C. Kraemer*
773 Effects of a Restricted Work-Site Smoking Policy on Employees Who Smoke *J. Brigham, J. Gross, M. L. Stitzer, and L. J. Felch*

DIET AND NUTRITION

779 Work-Site Cholesterol Screening and Dietary Intervention: The Staff Healthy Heart Project *A. Barratt, R. Reznik, L. Irwig, A. Cuff, J. M. Simpson, B. Oldenburg, J. Horvath, and D. Sullivan for the Steering Committee of The Staff Healthy Heart Project*
783 Improving Dietary Behavior: The Effectiveness of Tailored Messages in Primary Care Settings *M. K. Campbell, B. M. DeVellis, V. J. Strecher, A. S. Ammerman, R. F. DeVellis, and R. S. Sandler*

(continued on page 705)

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

American Public Health Association

May 1994, Vol. 84, No. 5 ISSN: 0090-0036 Established 1911

788 The Use of Vitamin Supplements and the Risk of Cataract among US Male Physicians *J. M. Seddon, W. G. Christen, J. E. Manson, F. S. LaMotte, R. J. Glynn, J. E. Buring, and C. H. Hennekens*

793 High Temporal, Geographic, and Income Variation in Body Mass Index among Adults in Brazil *R. Sichieri, D. C. Coitinho, M. M. Leão, E. Recine, and J. E. Everhart*

799 Dietary Sources of Fats and Cholesterol in US Children Aged 2 through 5 Years *F. E. Thompson and B. A. Dennison*

807 Overweight Schoolchildren in New York State: Prevalence and Characteristics *W. S. Wolfe, C. C. Campbell, E. A. Frongillo, Jr, J. D. Haas, and T. A. Melnik*

814 5-A-DAY: Dietary Behavior and the Fruit and Vegetable Intake of Latino Children *C. E. Basch, P. Zybert, and S. Shea*

PUBLIC HEALTH RESEARCH

819 The Fallacy of the Ecological Fallacy: The Potential Misuse of a Concept and the Consequences *S. Schwartz*

825 The Logic in Ecological: I. The Logic of Analysis *M. Susser*

830 The Logic in Ecological: II. The Logic of Design *M. Susser*

836 The Ecological Effects of Individual Exposures and Nonlinear Disease Dynamics in Populations *J. S. Koopman and I. M. Longini, Jr*

Public Health Briefs

VARIOUS TOPICS

843 Family Networks: Predictors of Nursing Home Entry *V. A. Freedman, L. F. Berkman, S. R. Rapp, and A. M. Ostfeld*

846 The Teaching of Occupational Health in US Medical Schools: Little Improvement in 9 Years *J. M. Burstein and B. S. Levy*

850 International Infant Mortality Rates: Bias from Reporting Differences *E. M. Howell and B. Blondel*

852 Urinary Biological Monitoring Markers of Anticancer Drug Exposure in Oncology Nurses *M. A. Newman, B. G. Valanis, R. S. Schoeny, and S. Q. Hee*

DIET AND NUTRITION

856 The Consumption of Well-Done Red Meat and the Risk of Colorectal Cancer *J. E. Muscat and E. L. Wynder*

859 Outbreak and Sporadic Egg-Associated Cases of *Salmonella enteritidis*: New York's Experience *D. L. Morse, G. S. Birkhead, J. Guardino, S. F. Kondracki, and J. J. Guzewich*

861 The Reproducibility of Data from a Food Frequency Questionnaire among Low-Income Latina Mothers and Their Children *C. E. Basch, S. Shea, and P. Zybert*

Other Journal Departments

706 APHA Masthead

709 Advertisers' Index

710 What *AJPH* Authors Should Know

713 Journal Masthead

865 Letters to the Editor

871 Book Corner

875 Job Opportunities

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 15th St, NW, Washington, DC 20005; (202) 789-5600; TDD (202) 789-5673. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*
Helen Rodriguez-Trias, MD, *Immediate Past President*
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Leonard M. Schuman, MD, *Vice-President (USA)*
John E. F. Hastings, MD, MPH, *Vice-President (Canada)*
Oscar Feo, MD, MSc, *Vice-President (Latin America)*
Jay H. Glasser, PhD, *Treasurer*
Allan G. Rosenfield, MD, *Chair, Executive Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*
Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*
Carol Easley Allan, PhD, RN (1997)
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Patricia P. Evans, MPH (1994)
Eugene Feingold, PhD, JD, *President*
Jay H. Glasser, PhD, *Treasurer*
Alan R. Hinman, MD, MPH (1995)
Margo D. Kerrigan, MPH (1995)
Joanne Lukomnik, MD (1996), *Vice-Chair*
Steven Uranga McKane, DMD, MPH (1995)
Carlos W. Molina, EdD (1997)
Linda Rae Murray, MD, MPH (1996)
Hila Richardson, DrPH, MPH (1996)
Helen Rodriguez-Trias, MD, *Immediate Past President*
Reed V. Tuckson, MD (1994)
Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*
Frances Kissling, *Chair, Action Board*
George E. Pickett, MD, MPH, *Speaker, Program Development Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*

American Journal of Public Health

The *American Journal of Public Health* (*AJPH*), published monthly, is the official Journal of the American Public Health Association, Inc, 1015 15th St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005. *AJPH* allotted $40; *The Nation's Health* allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters. Printed in the United States.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on the first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, *AJPH* Reprints, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the Journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The *American Journal of Public Health* is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

Chronic Neurological Sequelae to Organophosphate Pesticide Poisoning

Kyle Steenland, PhD, Barbara Jenkins, MA, Richard G. Ames, PhD, Michael O'Malley, MD, MPH, David Chrislip, BA, and John Russo, PhD

ABSTRACT

Objectives. This work was undertaken to determine whether there are any chronic neurological sequelae to acute organophosphate pesticide poisoning.

Methods. California surveillance data were used in a study of neurological function among 128 men poisoned by organophosphate pesticides in California from 1982 to 1990 and 90 referents. Tests included a neurological physical examination, 5 nerve conduction tests, 2 vibrotactile sensitivity tests, 10 neurobehavioral tests, and 1 postural sway test.

Results. After correcting for confounding, the poisoned group performed significantly worse than the referent group on two neurobehavioral tests (sustained visual attention and mood scales). When the data were restricted to men with documented cholinesterase inhibition (n = 83) or to men who had been hospitalized (n = 36), the poisoned subjects also showed significantly worse vibrotactile sensitivity of finger and toe. Significant trends of increased impairment were found with increased days of disability on a wide spectrum of tests of both central and peripheral nerve function.

Conclusions. While these findings are limited by low response rates and by small sample sizes for specific pesticides, this study was based on a large surveillance database and is the largest study to date of the chronic effects of organophosphate pesticide poisoning. The evidence of some long-term effects of poisoning is consistent with two prior studies. (*Am J Public Health.* 1994;84:731–736)

Introduction

Organophosphate pesticide poisoning results in well-known acute effects, which typically end in a few days. These effects are caused by the binding of cholinesterase and the buildup of acetylcholine. According to data assembled from poison control centers, there are approximately 10 000 cases of organophosphate pesticide poisonings annually in the United States.[1]

Some organophosphates may also cause a delayed-onset peripheral neuropathy several weeks after exposure, which primarily affects the extremities. Symptoms may persist or gradually decrease. This peripheral neuropathy results from severe inhibition of neuropathy target esterase, causing axonal degeneration.[2,3]

The literature is sparse as to whether an acute poisoning followed by apparent recovery has any subsequent chronic neurological effects. Two epidemiological studies with reasonable sample size and control series have indicated deficits among poisoned subjects on various neurobehavioral tests.[4,5] To investigate further the chronic sequelae of acute organophosphate poisoning, we conducted a study of 128 men aged 16 and over who were poisoned by an organophosphate pesticide in California between 1982 and 1990. We studied neurological function via neurobehavioral tests, nerve conduction tests, vibrotactile sensitivity tests, a test of postural sway, and a neurological physical examination.

Methods

Study Population

The state of California requires physicians to report known or suspected cases of organophosphate pesticide poisoning.[6] The pesticide illness reports generated by this system contain detailed information about patients' medical history and exposure; such information was collected from poisoned individuals and their employers during an investigation conducted by the Agriculture Department of the county where the poisoning occurred. Information is available regarding whether a patient was hospitalized as well as the number of days the patient took off from work (disability days) subsequent to the poisoning. In most cases, the report lists the specific pesticide thought to be responsible for the poisoning ("primary" pesticide), as well as any other pesticides to which the patient was exposed at the time of poisoning.

Case subjects (the "exposed" group) were the systemic organophosphate pesticide poisonings between 1982 and 1990 among men aged 16 and older (excluding suicide attempts). Women were not included owing to small numbers (15% of eligibles) and anticipated increased difficulty in tracing. All subjects had been exposed to one or more organophosphate pesticides and had sought medical attention (28% had been hospitalized for at least 1 night). Cases were divided into two types, "definite" and "probable." Definite

Kyle Steenland, Barbara Jenkins, David Chrislip, and John Russo are with the National Institute for Occupational Safety and Health, Cincinnati, Ohio. Richard G. Ames and Michael O'Malley are with the California Environmental Protection Agency in Berkeley and Sacramento, respectively.

Requests for reprints should be sent to Kyle Steenland, PhD, NIOSH R-13, 4676 Columbia Pkwy, Cincinnati, OH 45226.

This paper was accepted January 24, 1994.

TABLE 1—Comparison of Poisoned vs Nonpoisoned Subjects

	Poisoned (n = 128)	Nonpoisoned (n = 90)
Mean age, y (SD)	33.8 (9.5)	29.5 (10.9)*
Mean grade level (SD)	9.6 (4.9)	10.6 (3.6)
Mean body mass index (SD)	27.3 (4.4)	26.8 (4.7)
Race[a]		
Hispanic, %	56	56
White, %	41	39
Other, %	3	5
Preferred language[b]		
English, %	70	66
Spanish, %	30	34
Current smokers, %	37	33
Current drinkers, %	66	72
Mean hours of sleep, night before testing (SD)	7.2 (1.7)	7.2 (1.5)
Mean cups of coffee/tea on morning of test (SD)	0.8 (1.1)	0.9 (1.0)
Mean no. of drinks, night before testing (SD)	1.5 (2.7)	1.8 (2.4)
Average year of poisoning	1986	NA
Self-reported solvent exposure, %	20	21
Current employment in agriculture,[c] %	40	19*

Note. NA = not applicable.
[a]Self-reported.
[b]Language chosen for computerized neurobehavioral test; 18 subjects did not take this test.
[c]Nonpoisoned subjects were required to not be currently exposed to pesticides; nonpoisoned men in agriculture denied current exposure.
*Significant difference between groups at $P = .05$.

cases (n = 83) were those reporting one or more symptoms compatible with organophosphate poisoning (flulike symptoms, headaches, nausea, diarrhea, salivation), accompanied by a documented inhibition of red blood cell or plasma cholinesterase. Inhibition had to be at least 20% below the subject's own baseline value or be a value considered below the normal range by the testing laboratory. Probable cases (n = 45) did not have data on cholinesterase inhibition but reported either (1) compatible symptoms accompanied by relative specific physical signs—for example, bradycardia (heart rate less than 60) or miosis (narrowing of pupils to 3 mm or less); or (2) compatible symptoms accompanied by a history of direct exposure to the skin or eye during an application or spill. Virtually all subjects had been exposed occupationally.

Subjects were traced via the address in the pesticide illness report, telephone directory assistance, and current addresses obtained from the Internal Revenue Service. Social security numbers were not available. Subjects were contacted by phone and mail and asked to participate by coming to a central location in California, spending the night in a hotel, and taking a 4-hour battery of tests the following day. They were excluded if they had diabetes, had suffered a stroke, or had experienced trauma that had affected their central nervous system. Subjects were paid $100 for their participation and reimbursed for their expenses.

A nonexposed comparison group consisted of friends of the subjects not currently working with pesticides. The nonexposed individuals were also paid $100 for their participation. Participation was voluntary for all subjects, and consent forms were obtained.

Tests of Neurological Function

All tests were performed by technicians "blind" to exposure status.

Nerve conduction tests for peripheral neuropathy were done on three nerves of the dominant arm (median motor, median sensory, and ulnar sensory) and two nerves of the dominant leg (peroneal motor and sural sensory) using a standard protocol.[7,8] Two measures of nerve function were made: (1) conduction velocity—the amount of time from stimulation to the onset of principal depolarization (latency), divided by distance; and (2) peak amplitude—the size of the maximal response of the compound action potential for sensory studies or the M-wave for motor studies.

All tests were performed with surface electrodes; stimulation consisted of 100 to 200 microsecond square pulses. Motor responses were orthodromic; sensory tests were antidromic. Surface temperature of the limb was maintained at 33.0°C (centigrade) for the upper limb and 32.0°C for the lower limb.

Sensitivity to vibration was tested as a measure of the possible axonal degeneration in the sensory nerves of the index finger and big toe (dominant side, five trials). Vibrotactile thresholds were measured with the Vibration II instrument via the "method of limits."[9]

Eight computerized neurobehavioral tests from the Neurobehavioral Evaluation System (NES2, version 4.22) were conducted.[10] These tests were available in English and Spanish. While generally self-explanatory and requiring little intervention by the investigator, they did require a reading level equivalent to an elementary school education. The eight tests used are listed below:

1. Mood scales (affect test). Measures subjects' self-reported transient states of tension, depression, anxiety, fatigue, and confusion.
2. Finger tapping (motor speed test). Measures how many times a key can be repeatedly struck in 30 seconds, using the preferred hand.
3. Sustained visual attention (continuous performance test). Requires pressing a key quickly when a certain letter appears amid a temporal sequence of various letters; 60 letters appear in 5 minutes.
4. Hand–eye coordination (visuomotor accuracy test). Measures the degree of error in tracing a moving sine curve with a cursor.
5. Simple reaction time (visuomotor speed test). Measures how fast one can respond to a visual stimulus by pushing a button.
6. Symbol digit (coding speed test). Requires matching digits to symbols as fast as possible following an exhibited matched pattern.
7. Pattern memory (visual memory test). Requires selecting a previously seen pattern out of three similar patterns.
8. Serial digit learning (learning/memory test). Requires memorizing and replicating a series of eight digits as quickly as possible.

In addition, two noncomputerized neurobehavioral tests of psychomotor function—the Santa Ana dexterity test and the pursuit aiming test—from the World Health Organization core battery of tests were used.[11] These tests require turning rows of successive pegs 180 de-

TABLE 2—Adjusted Linear Regression Results on a Battery of Tests of Neurological Function for All Poisoned Subjects, Definitely Poisoned Subjects, and Hospitalized Subjects vs Nonpoisoned Control Subjects

Outcome	R^2 Model for All Poisoned[a]	All Poisoned (n = 128)			Definite Poisonings (n = 83)			Hospitalized Poisonings (n = 36)		
		Coefficient	Direction[b]	P[c]	Coefficient	Direction[b]	P[c]	Coefficient	Direction[b]	P[c]
NCV—median sensory	.15	−1.01	−		−1.06	−		0.26	+	
NCV—median motor	.07	−0.58	−		−1.19	−		−0.04	−	
NCV—ulnar sensory	.09	0.26	+		0.07	+		0.12	+	
NCV—peroneal motor	.19	−0.89	−	.09	−1.11	−	.06	−0.43	−	
NCV—sural sensory	.06	−0.04	−		0.41	+		0.22	+	
Amplitude—median sensory	.30	−1.55	−		−2.54	−	.10	−1.41	−	
Amplitude—median motor	.10	0.59	+		0.93	+	.06	1.63	+	.01
Amplitude—ulnar sensory	.12	−2.39	−		−1.65	−		−1.73	−	
Amplitude—peroneal motor	.03	−1.69	−	.06	−1.56	−		−1.35	−	
Amplitude—sural sensory	.21	−1.58	−		−2.04	−		−2.25	−	
Vibration—finger	.08	0.17	−		0.27	−	.03	0.65	−	<.01
Vibration—toe	.33	0.18	−		0.33	−	.05	0.60	−	<.01
Tapping	.18	−1.27	−		−4.89	−		−7.64	−	
Hand–eye coordination	.25	0.00	−		0.03	−		0.00	+	
Simple reaction time	.09	0.01	−		0.01	−		0.02	−	.10
Continuous performance	.13	0.01	−	.05	0.02	−	.01	0.03	−	.01
Symbol digit	.42	0.12	−		0.15	−		0.27	−	.04
Pattern memory	.08	0.78	+		0.63	+		0.18	+	
Serial digit	.27	−0.47	+	.10	−0.03	+		−0.51	+	
Mood scales[d]										
Tension	.03	0.24	−	.02	0.26	−	.03	−0.11	−	
Depression	.01	0.03	−		0.07	−		−0.08	+	
Anger	.02	0.04	−		−0.03	+		0.15	−	
Fatigue	.08	0.30	−		0.27	−		−0.08	+	
Confusion	.20	0.14	−	.01	0.14	−	.02	0.03	−	
Pursuit aiming	.08	0.13	+		0.93	+		−8.48	−	
Santa Ana dexterity	.21	0.53	+		1.21	+		−0.39	−	
Postural sway[e]	.03	0.03	−		0.03	−		0.05	−	

[a]All models had a variable for nonpoisoned vs poisoned. Models for nerve conduction velocity (NCV) and amplitude included age, race, and body mass index. Models for vibrometry included age, weight, race, and height. Models for computerized Neurobehavioral Evaluation System tests included age, education (grade), and test language (Spanish or English). Models for pursuit aiming and Santa Ana tests included race, age, and education. Models for postural sway included race, age, education, height, and weight.

[b]Coefficient represents difference in outcome between poisoned and nonpoisoned. Performance of poisoned vs referents, + = better; − = worse.

[c]*P* values are for tests of the null hypothesis that the coefficient equals 0. The actual values of the outcome variables, adjusted for covariates, may be obtained by request from the authors.

[c]*P* values reported if ≤.10.

[d]For mood scales, a higher score meant more tension, depression, etc., interpreted here as a worse performance.

[e]The difference in sway for eyes shut vs eyes open (higher scores represent worse performance).

grees and using a pencil to mark a point inside each of a series of circles as quickly as possible.

A computerized measurement of postural sway (30 seconds), a quantitative analogue of the Romberg clinical exam, was conducted to measure central nervous system function.[12] The outcome is the difference in sway between measurements taken with eyes open and eyes closed.

Finally, a physician conducted a standard neurological physical examination designed to detect gross neurological abnormalities and asked the subject about past medical history. The exam, which screened for primarily motor and cerebellar function, was used primarily to exclude subjects who suffered neuropathies from causes unrelated to pesticides, such as trauma or stroke.

Analysis

Thirty-eight poisoned subjects (30%) did not bring a friend to serve as a nonexposed comparison. Data were analyzed without reference to pair matching both to avoid loss of poisoned cases in the analyses and to facilitate multivariate analyses permitting control of unmatched confounders. All outcome measures were continuous variables; multivariate linear regression was used. Among the potential confounders considered were age, race (self-reported), body mass index (weight in kilograms/height in meters squared), education (grade level), preferred language, hours of sleep and alcohol consumption the evening prior to the test, smoking habits, use of prescription medicine, medical history, current exposure to solvents or pesticides (self-reported), and coffee consumption the morning of the tests. Multivariate models for nerve conduction velocities and amplitudes included race, age, and body mass index. Regressions for computerized neurobehavioral tests included age, grade level, and language in which the test was taken (Spanish or English); those for vibrotactile sensitivity and postural sway included age, race, weight, and height. All models were checked for assumptions of normality of residuals and model fit.

A dichotomous variable indicating poisoned case or referent was the principal exposure variable included in most

TABLE 3—Neurological Outcomes as a Function of Increased Days Taken Off from Work after Poisoning[a]

Outcome	R^2 for Model	Direction of Effect with Increased Days Taken Off from Work[b]	P[c]
NCV—median sensory	.16	+	
NCV—median motor	.08	−	
NCV—ulnar sensory	.08	+	
NCV—peroneal motor	.20	−	.03
NCV—sural sensory	.08	+	
Amplitude—median sensory	.33	−	.06
Amplitude—median motor	.10	+	
Amplitude—ulnar sensory	.13	−	
Amplitude—peroneal motor	.01	−	
Amplitude—sural sensory	.22	−	
Vibration—finger	.11	−	< .01
Vibration—toe	.34	−	
Tapping	.21	−	.03
Hand–eye coordination	.26	−	
Simple reaction time	.11	−	.03
Continuous performance	.16	−	< .01
Symbol digit	.44	−	< .01
Pattern memory	.09	−	.09
Serial digit	.26	−	.03
Mood scales			
Tension	.01	−	
Depression	.01	−	
Anger	.02	−	
Fatigue	.05	−	.10
Confusion	.18	−	
Pursuit aiming	.10	−	.06
Santa Ana dexterity	.21	−	
Postural sway	.02	+	

Note. NCV = nerve conduction velocity.

[a]"Days taken off from work" was entered as a continuous variable in models, replacing the dichotomous (poisoned/nonpoisoned) exposure variable used in other models. There were 18 poisoned men excluded from the analysis because they were missing data for days taken off from work. Referents were assigned "0" for this variable.
Thirty eight poisoned men had 0 days taken off from work and were grouped with the referents; arbitrary assignment of 1 day off from work for these 38 did not change results. Models included days taken off from work and relevant confounders (see Table 2).

[b]Performance of those who took more days off from work: + = better; − = worse.

[c]P values are for tests of the null hypothesis of no trend in outcome with days of work, and are reported if $P \leq .10$.

TABLE 4—Number of Poisoned Cases, by Type of Pesticide[a]

Pesticide	Primary[b]	Some Exposure
Chlorpyrifos	10	17
Diazinon	11	19
Dimethoate	7	14
Demeton methyl/ oxydemeton methyl	3	21
Mevinphos	13	39
Parathion	5	8
Phosalone	20	20
Other—specified	19	NA
Undetermined	38	NA

Note. NA = not available.

[a]Taken from original pesticide illness report.

[b]For most poisoned cases, a primary pesticide suspected of causing the poisoning was listed in the record, but for 38 poisoned subjects no primary pesticide was specified. In addition to the primary pesticide, many men were also exposed concurrently to another organophosphate, which could possibly have contributed to the poisoning (here listed under "some exposure"). All subjects in the category "primary exposure" for a pesticide are included in the category "some exposure" for that pesticide.

models. Analyses were also conducted in which the dichotomous exposure variable was replaced by a variable for number of days taken off from work (disability) subsequent to poisoning, with 0 days assigned to referents.

Although there were numerous outcomes, no formal correction for multiple comparisons was made[13]; instead, statistically significant associations were evaluated for coherence and consistency.

A number of individuals were excluded for all or some tests. The first 16 individuals tested did not have this mood scales test on the computerized neurobehavioral battery since this test was added after the first round of testing. Eighteen subjects (13 poisoned subjects and 5 nonpoisoned subjects, $P = .15$) did not take the computerized neurobehavioral tests, most often because their reading ability was insufficient. Other subjects were excluded in lesser numbers from specific tests for reasons such as malfunction of equipment, trauma affecting the nerves, and frostbite.

Results

We tested 128 poisoned and 90 nonpoisoned subjects. The 128 poisoned men represented 31% of the target population of 416 potential participants. Of the remaining potential participants, 13% refused, 37% could not be located, and 19% could not be contacted. (Although addresses were provided, mailings elicited no response from this latter group, and no phone number was listed with directory assistance.) A comparison of the 128 poisoned participants with the 288 poisoned nonparticipants revealed that nonparticipants were slightly older, were more likely to be Hispanic, and took more days off from work following their poisoning (data not shown).

Table 1 shows that poisoned and nonpoisoned subjects were generally similar but that poisoned subjects were significantly older and somewhat less educated.

Table 2 compares all poisoned subjects, definitely poisoned subjects, and hospitalized subjects with the 90 nonpoisoned subjects. For all poisoned subjects, the only statistically significant differences from the nonexposed group (at the $P = .05$ level) occurred for the test of sustained visual attention (continuous performance) and for two mood scale tests. For the 83 subjects with definite poisonings (documented cholinesterase inhibition), vibrotactile sensitivity for both finger and toe was also significantly worse. For the 36 hospitalized subjects (1 or more nights), vibrotactile sensitivity, the sustained attention test, and the symbol digit test were significantly worse than they were for the referents.

Table 3 shows the results of analyses by days taken off from work (disability days), which ranged from 0 to 32 for poisoned subjects (mean = 3.8, SD = 5.5). Significant trends of increased impairment with increased days of disability were found for peroneal nerve conduction velocity, for finger vibrotactile sensitivity, and for 5 of 10 neurobehavioral tests.

Variables for current employment in jobs with potential pesticide exposure and for years of self-reported past exposure to pesticides were tested but were not generally associated with the outcomes, nor did their inclusion in the model alter results for the effect of poisoning.

Analyses of the data according to year of poisoning did not change results, indicating that observed positive associations were not restricted to those subjects who were poisoned more recently.

Table 4 indicates the distribution of poisonings by pesticide, based on the original pesticide illness reports. As previously noted, these reports list a primary pesticide involved in the poisoning, as well as other pesticides to which a subject was exposed that might have played a role in the poisoning. In some cases, no one pesticide was deemed primary.

Table 5 shows the results of pesticide-specific analysis. For primary pesticides, chlorpyrifos- and phosalone-poisoned subjects showed some decrement in peripheral nerve function.

Table 5 also presents results for subjects exposed to any of seven pesticides at the time of poisoning, regardless of whether that exposure was considered the primary cause of the poisoning. Poisoned men who had been working with demeton methyl (Metasystox) showed a variety of decrements compared with the nonexposed men. (Eighteen of 21 men exposed to demeton methyl or oxydemeton methyl had no primary pesticide listed.) Similar findings for mevinphos may be partly owing to the fact that one third of the men working with mevinphos were also working with demeton methyl (13/39). Findings were unremarkable when analyses were restricted to those whose primary pesticide was mevinphos.

TABLE 5—Significant Associations ($P \leq .05$) in Analyses of Neurological Function, by Primary Pesticide or by Some Exposure to a Pesticide[a]

Primary[b]	
Phosalone[c]	More tension ($P = .05$) and fatigue ($P < .01$) on mood scales, better Santa Ana dexterity test ($P = 05$), worse sural amplitude ($P = .02$)
Diazinon	No associations
Chlorpyrifos	Worse peroneal motor nerve conduction velocity ($P = .04$) and ulnar sensory amplitude ($P = .03$)
Mevinphos	No associations
Some exposure	
Parathion	No associations
Diazinon	More tension on mood scales ($P = .04$)
Chlorpyrifos	More tension on mood scales ($P = .02$, worse finger vibrotactile sensitivity for the definitely poisoned ($P = .003$, n = 8)
Dimethoate	More tension ($P = .02$) and fatigue ($P = .05$) on mood scales
Mevinphos	Worse finger vibrotactile sensitivity ($P = .01$), worse performance on sustained attention test ($P < .01$)
Demeton methyl[d]	Worse finger ($P < .01$) and toe ($P = .01$) vibrotactile sensitivity; worse sustained attention ($P = .03$), simple reaction ($P = .04$), and symbol digit ($P < .01$) tests; more tension ($P = .05$) on mood scales

[a]Models used were the same as those used for Table 3, with same referent group (n = 90). Information on pesticide exposure at time of poisoning came from the original pesticide illness report.
[b]Analyses were restricted to the four primary pesticides with the most cases (nine or more).
[c]Phosalone only occurred as a primary pesticide.
[d]Demeton methyl and oxydemeton methyl (Metasystox and Metasystox-R) combined.

TABLE 6—Comparison of Neurobehavioral Tests in Three Studies of Men Poisoned by Organophosphate Pesticides

	Present Study				
Test	All Poisoned Subjects (n = 128)	Hospitalized Poisonings (n = 36)	Trend with Days Taken Off from Work (n = 110)	Savage et al.[4] (n = 100)	Rosenstock et al.[5] (n = 36)
Tapping	–	–	–*	–*	–
Simple reaction time	–	–	–	NA	–
Sustained attention	–*	–*	–*	NA	–*
Symbol digit	–	–	–*	–*	–*
Pursuit aiming	+	–	–	NA	–*
Santa Ana dexterity	+	+	–	–*	–*

Note. + = better performance; – = worse performance. NA = not available.
*$P < .05$.

Discussion

For the neurobehavioral tests used in this study, we found significantly worse performance on 2 of 10 tests (mood scales and the sustained visual attention test) by poisoned subjects, worse performance on a third test (symbol digit) by hospitalized subjects, and significant trends of worse performance on 5 of 10 tests by those who took more days off from work after poisoning. Increased impairment for those with presumably more severe poisonings (those hospitalized and those who took more days off from work) tends to support the hypothesis that the observed deficits on some of the neurobehavioral tests may be causally related to past poisonings. Two prior studies have also found that poisoned men performed significantly worse on a broad spectrum of neurobehavioral tests.[4,5] Table 6 compares our findings with those of two previous reports for six such tests. Our findings may be viewed as reasonably consistent, although the prior studies found a broader spectrum of deficits among poisoned subjects than we did. As most of the poisoned subjects in these studies had been hospitalized, they may have represented more severe poisonings.

Regarding the peripheral nervous system, nerve conduction velocities and amplitudes were not significantly worse for poisoned men overall. For pesticide-specific analyses, there were some significant deficits in nerve conduction velocity

and/or amplitude observed among men poisoned by chlorpyrifos or phosalone. This is consistent with a case report of delayed-onset neuropathy following poisoning by chlorpyrifos.[14]

Vibrotactile sensitivity of finger and toe was significantly reduced among poisoned men with documented cholinesterase inhibition, among poisoned men who had been hospitalized, and among men who took more days off from work after poisoning. Pesticide-specific analyses suggest that demeton methyl was the pesticide most associated with the vibrotactile sensitivity deficit among the poisoned men. A case report of delayed-onset neuropathy following poisoning by demeton methyl has been previously published.[15] A finding of vibrotactile sensitivity deficit in the absence of significant slowing of nerve conductions may reflect the relative insensitivity of electrophysiologic measures for the detection of a distal axonopathy.[16]

There are a number of limitations in our data. Our relatively low response rate may have affected our results in unpredictable ways. We anticipated difficulty in locating poisoned subjects, a number of whom were migrant Mexican workers. Possible conservative biases could have occurred if those subjects who did not participate were more likely to have had neurological impairment, but the opposite scenario is also possible. The use of friend controls could have also biased our findings toward the null if friends were selected for this similarity to the poisoned subjects regarding the outcomes under study (overmatching); however, a bias in the opposite direction could occur if the friends are more sociable people with better neurobehavioral capability.

Another limitation is the low proportion of variation explained by our predictor variables, indicating that individual heterogeneity or other environmental factors that we did not measure played a large role in determining the neurological outcomes of interest. It is possible that other unmeasured variables may have acted as confounders in our data. It should be noted that other studies using these tests have shown a similar low proportion of variation explained. A final limitation is the small sample size for pesticide-specific analyses, limiting our power to detect significant results.

Despite these limitations, our study is the largest to date of neurological function among both subjects previously poisoned by organophosphate pesticides and appropriate referents. It has the additional advantage of being population based, with a strict definition of cases.

In conclusion, we did not find apparent symptomatic damage to the neurological function of men poisoned in the past by organophosphate pesticides, but we did find some evidence of injury to the peripheral nerves as reflected by decreased vibrotactile sensitivity. We also found some evidence of deficits in central nervous system function as reflected by worse performance on 2 of 10 neurobehavioral tests. Furthermore, performance on a number of neurobehavioral tests was significantly worse for those with more severe poisonings, with severity measured by either hospitalization or number of days taken off from work subsequent to poisoning. □

Acknowledgments

This study was conducted as a collaborative effort between the State of California and the National Institute for Occupational Safety and Health (NIOSH). Financial support was provided partly by the US Environmental Protection Agency.

Drs Joseph Arezzo, Ann Fidler, and Jerry Blondell offered helpful comments on the manuscript. Numerous NIOSH employees contributed to this study: Dr Ray Alderfer, Dr Geoffrey Calvert, Dr Tony Suruda, Karen Davis-King, Nina Baird, Steve Brightwell, Robert Dick, Marie Haring-Sweeney, Kathy Watkins, Chris Gersic, Jean Geiman, Bettie Walpole, Bill Ehling, Rose Watkins, and Patty Gudlewski. Aurelio and Rosa Marta de los Santos recruited subjects and translated. Laura Weiss (Cal EPA) helped collect data. Nada Elmore, Jean Wilhite, and Pat Koenig (Biotrax) performed nerve conduction tests.

References

1. Litovitz T, Bailey K, Schmitz E, et al. 1990 annual report of the American Association of Poison Control Centers National Data Collection System. *Am J Emerg Med.* 1991;9:461–509.
2. Cherniak M. Toxicological screening for organophosphorus-induced delayed neurotoxicity. *Neurotoxicol.* 1988;9(2):249–272.
3. Kaloyanova F, El Batawi M. *Human Toxicology of Pesticides.* Boca Raton, Fla: CRC Press; 1991.
4. Savage E, Keefe T, Mounce L, et al. Chronic neurological sequelae of acute organophosphate pesticide poisoning. *Arch Environ Health.* 1988;43:38–45.
5. Rosenstock L, Keifer M, Daniell W, et al. Chronic central nervous system effects of acute organophosphate pesticide intoxication. *Lancet.* 1991;338:223–227.
6. Brown S, Ames R, Mengle D. Occupational illnesses from cholinesterase-inhibiting pesticides among agricultural applicators in California, 1982–1985. *Arch Environ Health.* 1989;44:34–39.
7. Kimura J. *Electrodiagnosis in Diseases of Nerve and Muscle: Principles and Practice.* Philadelphia, Pa: Davis; 1983.
8. Sweeney MH, Fingerhut M, Arezzo J, et al. Peripheral neuropathy after occupational exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Am J Ind Med.* 1993;23:845–858.
9. Gerr F, Hershman D, Letz R. Vibrotactile threshold measurement for detecting neurotoxicity. *Arch Environ Health.* 1990;45: 148–154.
10. Letz R. *Neurobehavioral Evaluation System 2 (NES2): User's Manual.* Winchester, Mass: Neurobehavioral Systems, Inc; 1988.
11. Johnson B, ed. *Prevention of Neurotoxic Illness in Working Populations.* New York, NY: John Wiley & Sons Inc; 1987.
12. Dick R, Bhattacharya A, Shukla R. Use of a computerized postural sway measurement system for neurobehavioral toxicology. *Neurotoxicol Teratol.* 1990;12:1–6.
13. Thompson W. Statistical criteria in the interpretation of epidemiologic data. *Am J Public Health.* 1987;77:191–194.
14. Lotti M, Moretto A, Zoppellari R, et al. Inhibition of lymphocytic neuropathy target esterase predicts the development of organophosphate-induced delayed neuropathy. *Arch Toxicol.* 1986;59:176–179.
15. Fournier E. Preuves clinic et biologiques des intoxications par produits de maison, Rapport de seance pleniere UPAC, Prague 1968. *Pure Appl Chem.* 1969;18:151.
16. Arezzo J, Schaumburg H. Office and field diagnosis of neurotoxic disease. *J Am Coll Toxicol.* 1989;8:311–319.

Health Status Differentials in the People's Republic of China

James Sutherland Lawson, MD, MHA, and Vivian Lin, DrPH

ABSTRACT

Objectives. This study sought to demonstrate that health status varies markedly in different parts of China.

Methods. The main source of data was statistics compiled by the Chinese Ministry for Public Health for 1978 to 1990 regarding causes of death. However, because mortality statistics in China are based on localities that have the capacity to provide data, they are not entirely representative. The international classification of disease categories was also used, together with anatomically based disease descriptions. Rates were calculated using the 1982 and 1990 population censuses.

Results. Death rates differ markedly between urban and rural areas. Deaths due to infectious diseases, respiratory diseases, pregnancy and childbirth, and injuries and poisoning are much higher in rural areas; those due to pertussis, dysentery, typhoid, hepatitis, rabies, and anthrax are much more common in the apparently poorer provinces. Schistosomiasis remains a major problem in some provinces. Goiter and cretinism are still major diseases in many parts of China, especially those areas with iodine deficiency.

Conclusions. Cause-of-death patterns in Chinese cities are similar to those of industrially developed countries such as Australia, Japan, and the United States. Such patterns in the poorer rural areas are much more typical of those of developing countries. (*Am J Public Health.* 1994; 84:737–741)

Introduction

Since the 1949 revolution and the consequent establishment of the People's Republic of China, the improved health status of the Chinese people has been dramatic. The significant achievements in this regard, both nationally and in some specific localities such as Shanghai,[1,2] have been widely acknowledged and documented.[3,4] Those achievements have been measured by increased life expectancy, decreased infant mortality, and reduced morbidity. Specifically, life expectancy increased from about 35 to 70 years between 1949 and 1990, and the overall death rate declined from approximately 20% to 6% of the population per year during the same period.[5] This decline in deaths appears to be primarily owing to the reduction in infectious diseases.[6]

Despite these improvements, health status varies markedly in different parts of China. Deaths due to all major causes are from 10% to 100% higher for specific age groups in rural as compared with city populations. In addition, deaths due to infections remain high in many provinces. These variations may be explained by the substantial differences in both geography and socioeconomic development across China.

Methods

Sources of Data

For the most part, data used in this paper are from statistics compiled by the Chinese Ministry of Public Health for 1978 to 1990.[7] Death rates are reported for gender and age groups by *International Classification of Disease* (ICD) categories and are based on the 1982 and 1990 population censuses and a 1% sample survey of the population in 1987.

A number of points need to be made about the use and interpretation of these data. First, informants in the Ministry of Public Health suggest that there may be some differences in diagnostic and classification approaches in China despite the use of ICD codes. Accordingly, we have presented selected tables as published in the original Chinese form. This selection of data is based on both the need to simplify a large volume of material and the interest in women of childbearing age.

Second, local quality control in the collection and reporting of statistics is likely to vary, and the localities included are probably those that are above average in their capacity to collect data. Thus, the data probably come from those areas of China with better health status. Nonetheless, these are the most reliable and comprehensive data currently available to non-Chinese persons.

Geographical Classifications

The geographical classifications are as follows:

1. Large cities: 16 cities, including Beijing, Tianjin, Shanghai, Nanjing, Guangzhou, Xian, and Wuhan;
2. Medium and small cities: 25 cities, including Suzhou, Hefei, Xiamen, Guilin, and Taiyuan;
3. Class I rural areas: 35 localities, including rural suburbs of Beijing, Tianjin, and Shanghai, and se-

James Sutherland Lawson is with the School of Health Services Management, University of New South Wales, Kensington, Australia. Vivian Lin is with the Department of Health and Community Services, Victoria, Australia.

Requests for reprints should be sent to James S. Sutherland, MD, MHA, School of Health Services Management, University of New South Wales, PO Box 1, Kensington NSW 2033 Australia.

TABLE 1—Death Rates per 100 000 Chinese Men, 1988

Causes of Death	Urban		Rural		
	Large Cities	Medium and Small Cities	Class I	Class II	Class III
Infectious diseases	19	26	29	49	60
Neoplasms	150	115	165	77	64
Endocrine, nutrition, and immune disorders	7	5	3	5	8
Blood disorders	1	1	1	1	1
Mental disorders	5	5	5	3	2
Central nervous system disorders	6	3	4	4	3
Heart disease	87	46	90	61	77
Cerebrovascular diseases	130	82	119	93	64
Respiratory system diseases	83	97	151	174	182
Digestive diseases	26	29	29	45	62
Genitourinary system diseases	10	8	10	9	12
Congenital anomalies	6	6	8	4	5
Perinatal conditions	750	710	1023	776	1423
Injuries and poisoning	46	55	71	93	103
Total	605	501	719	664	687

Source. Compiled from data[6] from the Chinese Ministry of Public Health, with permission.

TABLE 2—Death Rates per 100 000 Chinese Women, 1988

Causes of Death	Urban		Rural		
	Large Cities	Medium and Small Cities	Class I	Class II	Class III
Infectious diseases	11	20	18	38	48
Neoplasms	107	78	103	58	43
Endocrine, nutrition, and immune disorders	13	7	5	6	9
Blood disorders	2	2	1	2	2
Mental disorders	7	6	7	3	2
Central nervous system disorders	5	3	3	3	2
Heart disease	105	57	95	55	95
Cerebrovascular diseases	128	82	122	84	54
Respiratory system diseases	87	97	138	176	201
Digestive diseases	22	24	23	34	48
Genitourinary system diseases	10	7	7	7	10
Complications of pregnancy and childbirth	19	40	43	88	226
Congenital anomalies	6	6	7	4	3
Perinatal conditions	574	539	813	618	1243
Injuries and poisoning	39	46	51	77	66
Total	576	462	615	608	631

Source. Compiled from data[6] from the Chinese Ministry of Public Health, with permission.

lected cities and counties of Zhejiang and Jiangsu provinces;

4. Class II rural areas: 35 counties in Jiangxi, Hubei, Hunan, Guangdong, Fujian, and Sichuan provinces;
5. Class III rural areas: 9 counties selected from Guizhou, Anhui, and Gansu provinces.

Formal statistics about levels of income and education with respect to these geographic classifications were not available. However, it is obvious from firsthand observation that these levels are much higher in the large cities and class I rural areas than they are in medium and small cities and particularly in class II and III rural areas. At one extreme, the larger cities are the commercial, industrial, administrative, and cultural centers of China; at the other extreme, the provinces from which the class III areas are chosen are the poorest, most backward provinces in China. To offer some perspective, data from Australia, the United States, Japan, and India are also presented.

Results

Overall Death Rates

The 1988 age-adjusted death rates (for men and women) for the urban and rural areas by ICD codes are shown in Tables 1, 2, and 3. However, although the Chinese authorities use ICD codes, they do so with "Chinese characteristics"—namely, the frequent and continued use of anatomical classifications—which sometimes precludes meaningful international comparisons.

Death rates between urban and rural areas differ markedly. Of particular note are those differences in infectious diseases, respiratory diseases, digestive diseases, problems of pregnancy and childbirth, conditions of the perinatal period, and injuries and poisoning. While men are at greater risk of death from infectious diseases, malignant neoplasms, conditions of the perinatal period, and injuries and poisoning, women appear to be at slightly greater risk of death from endocrine, nutritional, and immune disorders and heart disease.

The most striking differences within urban areas are the higher death rates in small and medium cities due to pregnancy and childbirth, infectious diseases, respiratory diseases, and injuries and poisoning, and the higher death rates in larger cities due to neoplasms, conditions of the perinatal period, heart disease, and cerebrovascular diseases. The male/female differentials remain similar to the overall picture.

The differences in death rates between rural areas are dramatic. The gradient from class I to class III is steeply upward for infectious diseases, respiratory diseases, digestive diseases, pregnancy and childbirth among women, and, among men, injury and poisoning. There is also a reverse gradient in rural areas from class I to class III for malignant neoplasms and cerebrovascular diseases. The differential between different classes of rural areas is greater than that between large and medium/small cities.

The death rate for China as a whole was 6.28% for 1990 (calculated from the 1990 census). However, this rate varied

considerably between provinces and autonomous regions. For example, death rates were 5.4% in Beijing, 5.3% in Guangdong, 6.4% in Shanghai, 7.1% in Sichuan, 7.7% in Yunnan, and 9.2% in Tibet. These death rates (which are not age standardized) tend to reflect differing levels of socioeconomic development in specific areas.

The death rate due to infectious diseases is at historically low levels in all parts of China (1.2 per 100 000 in 1990). However, there are considerable regional variations, with deaths due to pertussis, dysentery, typhoid, hepatitis, rabies, and anthrax being much more common in poorer provinces such as Guizhou and Tibet than elsewhere in China. Schistosomiasis remains a major problem with 400 326 cases detected in 1989, more than half of which were in Hubei Province.

Goiter and cretinism are still major diseases in many parts of China. In 1989 there were 6 054 000 known patients with goitrous disease and 199 033 patients with cretinism. These conditions occur in the iodine-deficient regions, particularly in mountainous areas such as Xinjiang, Zhejiang, Fujian, Hubei, Guangxi, and Qinghai.

Deaths associated with pregnancy and childbirth appear to be high (compared with those in industrially developed countries). Based on surveys of such deaths conducted in 1989, maternal mortality rates per 100 000 females per year were 49.9 in urban areas and 114.9 in rural areas. Pre- and postpartum hemorrhage, hypertension, heart disease, pueperal fever, and amniotic embolism were the most frequent causes of maternal death.

International Comparisons

As shown in Table 4, there are several major differences in patterns of causes of death between China, the United States, Australia, and Japan. Deaths due to infectious diseases (particularly tuberculosis), stroke, bronchitis and emphysema, accidents, suicide among women, and maternity are higher in China; those due to cancer are lower in China than in Japan, the United States, and Australia.

As shown in Table 5, there are major differences in causes of death between China and India. Whereas the available data use anatomical classifications and are not age standardized, the pattern of deaths shown for India remains typical of that for a developing country, whereas that for China reflects a "transitional" stage in the rural areas.

TABLE 3—Leading Causes of Death in China (Death Rates per 100 000) by Selected Age/Gender Groups, 1988

Causes of Death	Urban		Rural		
	Large Cities	Medium and Small Cities	Class I	Class II	Class III
Males aged <1 y					
Perinatal conditions	747	710	1024	776	1424
Congenital anomalies	252	185	345	125	175
Respiratory system diseases	192	323	427	838	2098
Injuries and poisoning	44	120	233	229	582
Digestive diseases	40	72	52	140	213
Infectious diseases	32	61	59	83	225
Endocrine, nutrition, and immune disorders	...	...	...	...	175
Females aged 25 to 30 y					
Injury and poisoning	16	26	33	66	49
Neoplasms	10	11	10	10	15
Circulatory diseases	6	10	9	16	15
Infectious diseases	4	5	4	13	23
Central nervous system disorders	2	...	...	...	...
Complications of pregnancy and childbirth	...	4	5	11	29
Females aged 60 to 65 y					
Circulatory diseases	545	375	443	575	569
Neoplasms	377	365	379	310	220
Respiratory system diseases	301	237	236	512	475
Digestive diseases	114	55	166	124	129
Injuries and poisoning	42	65	63	126	75
Endocrine, nutrition, and immune disorders	40	...	...	...	...
Infectious diseases	...	48	34	128	94

Source. Compiled from data[8] from the Chinese Ministry of Public Health, with permission.

TABLE 4—Death Rates per 100 000 Male and Female Population in China, Japan, the United States, and Australia, by Selected Causes, 1989 to 1990

Causes of Death	China		Japan		United States		Australia	
	Males	Females	Males	Females	Males	Females	Males	Females
Infectious diseases (approximately 50% due to tuberculosis)	31	21	9	4	10	7	4	3
Cancer	129	86	150	76	183	110	164	102
Circulatory diseases (mainly ischemic heart disease)	197	199	165	108	283	159	267	158
Cerebrovascular diseases	112	106	64	44	34	30	47	42
Bronchitis/ emphysema	98	93	11	4	8	5	14	7
Maternal deaths		1.6		...		...		...
Accidents	65	51	51	21	81	28	67	25
Suicide	15	20	16	9	17	4	19	5
Total	634	565	521	294	715	423	649	378

Note. The death rates for Japan, the United States, and Australia have been age standardized by using a standard "world" population age distribution. The death rates for China are not age standardized.

Source. Data are adapted from the *World Health Statistics Annual*[9] and the Chinese Ministry of Public Health,[8] with permission.

TABLE 5—Deaths (%) by Major Cause Groups in Urban and Rural China and India

Causes of Death	China		India	
	Urban	Rural	Urban	Rural
Senility	...	...	NA	23.8
Malignant tumors	21.9	17.5	4.4	NA
Cerebrovascular diseases	20.8	16.2	NA	NA
Circulatory diseases	NA	NA	20.2	10.9
Respiratory system diseases	15.8	24.8	8.9	20.2
Heart disease	15.8	10.8	NA	NA
Injury and poisoning	6.9	10.7	10.5	6.6
Digestive diseases	4.0	5.0	5.7	6.6
Newborn diseases	1.5	2.5	12.2	9.8
Genitourinary system diseases	1.6	1.5	...	...
Endocrine, nutrition, and immune disorders	1.7	...	3.1	...
Infectious diseases	...	1.8	18.8	7.4
Tuberculosis	1.2	1.9	NA	NA
Maternal mortality	10	12	NA	20–30
Infant mortality	14.0	21.0	NA	80.0

Note. Data for urban India are for 1986; all remaining data are for 1989 to 1990. For both countries, maternal mortality rates are per 10 000 life births; infant mortality rates are per 1000 live births. NA = not available.

Source. Compiled from data from the Chinese Ministry of Public Health[9] and the *Bulletin of Regional Health Information*,[10] with permission.

Discussion

Despite reservations about the accuracy of the data, particularly as the most reliable information inevitably has come from China's more developed cities and provinces, the patterns of disease and injury as shown in the tables are a good reflection of firsthand observation. Traffic crashes are commonly seen; unsafe water, unsafe occupational health and safety practices, and unsafe food preparation are obvious to all visitors. In addition and most importantly, the nationwide experience has been confirmed in detail by epidemiological field studies conducted in the Shanghai area in the early 1980s.[2]

This review of mortality data clearly demonstrates that, despite dramatic improvements in the health status of the Chinese people during the past 40 years, a substantial burden of sickness and premature death remains. This burden does not affect the population equally. Deaths among infants due to respiratory infections, diarrheal diseases, and accidents are up to 10 times more common in poor rural communities than in large city populations. Maternal deaths are high by international standards and are also much more common in poor rural areas. Deaths due to infectious diseases generally, as well as to tuberculosis and respiratory infections in particular, are high for all of China but are much higher in the rural areas.

The incidence of trauma is also massive, with approximately 500 000 deaths per year; this has recently been addressed by Li and Baker.[11] Because of the vast population, the number of daily deaths due to traffic crashes and other trauma in China is equivalent to two fully laden jumbo jets crashing each day!

In both city and urban populations, coronary heart disease, cerebrovascular disease, and cancer are the most common causes of death among 65- to 70-year-old persons. A longitudinal study of adults in Shanghai by Chen et al.[12] has confirmed these trends. Unfortunately, and again because of the vast population, these deaths tend to hide the continuing problems of infectious diseases, trauma, and maternal deaths outlined above.

Tobacco consumption is rapidly rising in China, in parallel with a rapid rise in lung cancer. This issue has recently been addressed in detail by Yu et al.,[13] who regard the increased Chinese tobacco consumption as an "impending catastrophe."

Apart from maternal health programs, women's health issues have largely been unrecognized in China. Of particular significance and sadness are the reports of high female infant death rates owing to drowning, mechanical suffocation, and homicide. This may provide evidence to support anecdotes about female infanticide, especially in light of China's population policy. At later ages, the high rate of suicide suggests that the quality of life for women, especially given the social pressures confronting them, still means that "women's liberation" is more rhetoric than reality.

The current health status of the Chinese people largely reflects the current social and economic development in different parts of China. The much higher health status of Chinese urban dwellers can be accounted for by the higher education, more regular and varied food supplies, better sewerage, better housing and water supplies, better developed medical services, and better access to health services than is found in poor rural areas. Indeed, statistics about hospital beds, immunization, and running water all reflect similar differentials.[6] For instance, in the largest cities of Beijing and Shanghai, there are more than 4.5 hospital beds per 1000 population; immunization levels among 18- month-old infants is well over 90%; and more than 80% of these populations, including 50% of the nearby rural populations, have running water. In the relatively developed provinces of Jiangsu, Fujian, Hunan, and Yunnan, the number of hospital beds per 1000 population is above two, the immunization level is more than 90%, and 40% to 80% of the population (and 20% of the rural population) have access to running water. However, in the least developed provinces such as Guizhou, Gansu, and Anhui, the number of hospital beds per 1000 population is below two, immunization levels are below 80%, and only 20% to 50% of the population (and 10% of rural population) have access to running water.

In the 1980s, the Chinese health care system underwent significant changes as part of the overall economic reforms.[14,15] These changes may have exacerbated the wealth differential between the larger coastal cities and the hinterland. As the larger coastal cities have become "open," there has been a rapid increase in their standard of living and health care investments. Additionally, changes in agricultural policy and the rural economy have meant significant improvements in the standard of living for those areas close to the large cities. Many class I rural areas during the decade became absorbed into (or contiguous with) the urban communities; hence, large cities and rural class I share many similarities in death rates.

From the perspective of health status, China as a whole can be regarded as in transition, halfway between being an economically developed, industrialized country and being a poor developing country. Rural China is currently experiencing rates of infectious diseases, maternal deaths, and accidents typical of India and Africa, while urban China is experiencing rates of heart disease, stroke, and cancer typical of the United States, Japan, and Australia. In summary, the patterns of disease in urban China are typical of more developed countries and those patterns in rural China remain typical of developing countries.

The massive issue of population growth, although not considered in this article, is of major relevance to China's economic development and health status. If current trends continue, China's population is expected to rise from 1 200 000 000 in 1991 to approximately 1 500 000 000 by 2025, at which point the level might stabilize.[16] This growth in population implies a need to house, feed, and provide services for an additional population equivalent to that of the United States during the next 30 years.

The Chinese leadership has developed health policies[17] that seek to address these problems. These policies include improving the quality of drinking water, the techniques of birth, the hygiene of food supplies, and health services management; increasing hospital services and the number of medical graduates; and focusing on the prevention of infectious diseases.

The data presented in this paper suggest that several additional priorities could be addressed. These are (1) the need for vigorous anti-tobacco policies to reduce the problems of lung cancer and respiratory disease, (2) enhanced trauma prevention activities, and (3) development of programs aimed specifically at women's health issues. □

Acknowledgments

The assistance offered by the Ministry of Public Health, People's Republic of China, in providing the data for this paper is acknowledged with gratitude.

References

1. *China: The Health Sector.* Washington DC: World Bank; 1989.
2. Health services in Shanghai county. *Am J Public Health.* 1982;72(suppl).
3. Sidel V, Sidel R. *Serve the People.* Boston, Mass: Beacon Press; 1973.
4. Sidel R, Sidel V. *The Health of China.* Boston, Mass: Beacon Press; 1982.
5. *Chinese Health Statistical Digest, 1990.* People's Republic of China: Ministry of Public Health; 1991.
6. *Health Statistics in China.* People's Republic of China: Ministry of Public Health; 1988.
7. *Selected Edition on Health Statistics of China 1978–1990.* People's Republic of China: Ministry of Public Health; 1991.
8. *World Health Statistics Annual.* Geneva, Switzerland: World Health Organization; 1992: 310–315.
9. *Chinese Health Statistics 1989–1990.* People's Republic of China: Ministry of Public Health; 1990.
10. *Bulletin of Regional Health Information 1988–1990.* New Delhi, India: World Health Organization; 1992;166 (tables 4a, 4b).
11. Li G, Baker SP. A comparison of injury death rates in China and the United States 1986. *Am J Public Health.* 1991;81:605–609.
12. Chen Z, Peto R, Collins R, et al. Serum cholesterol concentration and coronary heart disease in populations with low cholesterol concentrations. *Br Med J.* 1991; 303: 276–282.
13. Yu JJ, Mattson ME, Boyd GM, et al. A comparison of smoking patterns in the People's Republic of China with the United States. *JAMA.* 1991;264:1575–1579.
14. Hsiao W. Transformation of health care in China. *N Engl J Med.* 1984;310(14):932–936.
15. Yang P-L, Lin V, Lawson JS. Health policy reform in People's Republic of China. *Int J Health Serv.* 1991;21(3):481–491.
16. *State of World Population.* New York, NY: United Nations; 1991.
17. Cui YL, ed. *Public Health in the Peoples Republic of China 1986.* Beijing, China: People's Medical Publishing House; 1987.

Pap Smear and Mammogram Screening in Mexican-American Women: The Effects of Acculturation

Lucina Suarez, MS

ABSTRACT

Objectives. For Mexican Americans, acculturation is a multidimensional process of adopting attitudes, values, and behavior from the non-Hispanic White culture. This study examines the effects of different dimensions of acculturation on the cancer screening behavior of Mexican-American women.

Methods. Subjects were 450 randomly selected Mexican-American women age 40 years and older living in El Paso, Texas. Personal interviews solicited information on age, income, education, health insurance, Pap smear and mammogram use, and acculturation. Acculturation was measured with five scales that assessed English proficiency, English use, value placed on culture, traditional family attitudes, and social interaction.

Results. The 2-year prevalence of Pap smear and mammogram screening increased with each gain in acculturation on English proficiency and use. These associations disappeared when adjusted for age, income, insurance, and education. After adjusting for sociodemographic factors and other acculturation dimensions, a strong traditional Mexican attitude toward family was positively related to mammography use.

Conclusions. Taking advantage of the positive influence of Hispanic familism on cancer screening behavior may increase the effectiveness of cancer control interventions in Mexican Americans. (*Am J Public Health.* 1994;84:742–746)

Introduction

Participation in breast and cervical cancer screening programs among Hispanic women must be improved. Of the three major ethnic and race groups in the United States (Non-Hispanic Whites, Blacks, and Hispanics), Hispanic women are the least likely to utilize Pap smears and mammograms.[1–3] The 1987 National Health Survey showed that Hispanic women age 18 years and older were twice as likely as Black women and three times as likely as White women to have never had a Pap smear.[1] Among women age 40 years and older, fewer Hispanic women (16%) than Black women (18%) or non-Hispanic White women (24%) have had a mammogram in the prior 3 years. Of the three major US Hispanic subgroups (Cuban, Puerto Rican, Mexican American), Mexican Americans are the least likely to use preventive health services in general.[4] Most studies among the Mexican-American population show that the lower rates of cancer screening are primarily associated with lower education levels, lack of health insurance coverage, and lower income levels.[2–5] Apart from these socioeconomic factors, others suggest that acculturation and assimilation levels of Mexican-American women are barriers to more frequent use of cancer screening tests (e.g., Pap smears and mammograms).[2,4]

Published studies examining factors other than socioeconomic factors have shown that high acculturation on English language use is associated with greater recency of Pap smear screening or mammography.[2,4,5] Only two of these studies have examined acculturation dimensions other than language. In a study of 603 Hispanic women age 55 years or older living in public housing projects in Los Angeles, in which four dimensions of acculturation (language preference, country of birth, contact with homeland, and attitude about children's friends) were examined, only language was associated with Pap smear and mammogram screening.[6] In a large-scale study based on the 1982–1984 Hispanic Health and Nutrition Examination Survey, acculturation measures were limited to language use (the only predictor of screening), ethnic identification, and generation status.[4] These studies confirm the importance of language but provide little information on the effect of cultural attitudes and values on the cancer screening behavior of Mexican-American women.

Acculturation—a process in which individuals whose primary learning has been in one culture (Mexican American) adopt attitudes, values, and behaviors from another culture (non-Hispanic White)—has many more dimensions than language use. English language use measures in some way functional integration into the non-Hispanic White society (i.e., assimilation), but it does not necessarily measure new attitudes or values. Country of origin or generational status of an individual by its unchanging nature fails to fully measure the continuous process of acculturation and assimilation. Depending on region of the country, neighborhood, and immigrant experience, a first-generation Hispanic American may be more acculturated and assimilated than a second- or third-generation Hispanic American.[7]

Lucina Suarez is with the Associateship of Disease Prevention, Texas Department of Health, Austin, Tex.

Requests for reprints should be sent to Lucina Suarez, MS, Associateship of Disease Prevention, Texas Department of Health, 1100 W 49th St, Austin, TX 78756.

This paper was accepted July 7, 1993.

The development of effective interventions for Mexican-American women requires understanding the effect of different dimensions of assimilation and acculturation on cancer screening behavior. The present study examines the influence of acculturation and assimilation on the cancer screening behavior of Mexican-American women living in a large Southwestern border community. In addition to structural assimilation, which includes both language use and social interaction with members of mainstream society, the effect on screening behavior of the two separate acculturation dimensions of cultural values and family attitudes was examined in light of other socioeconomic barriers.

Methods

The sample included 450 Mexican-American women age 40 years and older living in 16 census tracts along the United States-Mexico border in El Paso County. In 1990 the population within these 16 census tracts was at least 90% Hispanic and had a median income at or below the federal poverty threshold. All households in the study census tracts were identified by using the Polk city directory and the *1990 Census Tract City of El Paso Street Code Reference Manual.* A listing of street addresses by tract was used to randomly select 1300 households as a sampling frame to screen for Mexican-American women age 40 years and older. From this sampling frame with household enumeration, 549 households were identified as having at least one eligible woman for inclusion in the survey. Only one eligible woman from each household was interviewed. The decision rule for selecting from two or more eligible women was to interview the woman who had the most recent birthday because this information was on the enumeration form and was easy to determine. Of these households, 20 women were not available for an interview, 76 refused to give an interview, and 3 women gave incomplete interviews. Four hundred fifty women completed personal interviews for a response rate of 82.0%.

After determining whether the respondent preferred to be interviewed in Spanish or English, bilingual interviewers used a standard instrument to ask women their age, annual household income, education, health insurance coverage, Pap smear and mammogram screening practices, and questions related to structural assimilation and acculturation. Acculturation was measured by using previously published scales developed by Hazuda et al.[8] that assessed the following dimensions of assimilation and acculturation: adult proficiency in English, adult pattern of English vs Spanish language usage, value placed on preserving Mexican cultural origin, attitude toward traditional family structure and sex-role organization, and adult interaction with members of mainstream society. The scales have demonstrated construct validity, are applicable to Mexican-American populations, and provide measures of acculturation that are independent of each other.[8]

TABLE 1—Percentage Distribution of Mexican-American Women Age 40+ Years, by Acculturation Level, El Paso, Texas, 1991

		Acculturation Level			
	No.	Low 1	2	3	High 4
English proficiency	447	44.3	14.3	17.0	24.4
English vs Spanish usage	444	64.6	16.4	15.5	3.4
Value placed on cultural origin	442	27.6	35.3	26.0	11.1
Attitude toward traditional family	444	21.6	40.3	37.4	0.7
Interaction with mainstream society	444	56.1	30.4	12.6	0.9

Women were assigned a summary score on each dimension by adding response scores for items under each acculturation scale. Within each dimension, women were grouped into four levels rank-ordered from least to most acculturated based on previously defined cut points from the San Antonio Heart Study.[9] The use of these cut points, rigorously determined from a large population sample of Hispanics and non-Hispanic Whites, provides a standard method of measuring acculturation across different populations (e.g., El Paso and San Antonio). Location in the higher-numbered levels represented acculturation toward non-Hispanic White culture. For example, women in the highest level for proficiency in English were highly proficient in understanding, speaking, and writing English. Women in the highest level for English vs Spanish language pattern used mostly English when interacting with family, friends, neighbors, and coworkers, and they watched, listened to, and read English media. For value placed on preserving Mexican cultural origin, women in the highest level thought it was not important for children to know the history of Mexico, follow Mexican customs, or celebrate Mexican holidays. For attitude toward traditional family structure and sex-role organization, women in the highest level thought it was not important that family members know their family tree, have close ties to extended family, remember deceased family members, have married children live close to parents, or concede authority to father or brothers. For women in the highest level of interaction with members of mainstream society, the majority of their close friends, neighbors, and coworkers were non-Hispanic Whites.

Cancer screening history was ascertained from responses to the following questions: (1) Have you ever had a Pap smear (mammogram)? (2) When did you have your most recent Pap smear (mammogram)? Women were categorized according to whether or not they reported a Pap smear or mammogram in the previous 2 years. Pap smears or mammograms performed because of a health problem were considered diagnostic exams and were excluded. Thus, the outcome of interest was the use of screening Pap smears or screening mammograms as routine prevention services in the absence of symptoms.

Chi-square tests, including tests for trends, were performed to assess differences in prevalence of Pap smear and mammogram screening across various levels of sociodemographic factors and acculturation dimensions. With use of multiple logistic regression, odds ratios of having a screening Pap smear or screening mammogram within the previous 2 years were computed for each gain in acculturation level. Test-based 95% confidence intervals were computed with the software EGRET.[10] Univariate odds ratios and their 95% confidence intervals were computed for each acculturation dimension. Odds ratios and confidence intervals then were adjusted for relevant sociodemographic factors: age (10-year intervals), education (less than 7th grade, less than high school, high school graduate), income (don't know, <100% federal pov-

TABLE 2—Prevalence of Screening among Mexican-American Women Age 40+ Years, by Demographic Characteristics, El Paso, Texas, 1991

	Pap Smear within 2 y (n = 420),[a] %	Mammogram within 2 y (n = 416),[a] %
Age, y	*	
40–49	61.9	16.8
50–59	44.8	23.4
60–69	43.3	23.3
70+	28.1	22.2
Education	*	
<7th grade	40.9	17.8
<high school	51.8	25.3
High school graduate	59.2	27.0
Income	*	*
Don't know	38.1	21.4
<100% poverty level	46.2	18.7
<200% poverty level	51.6	27.3
>200% poverty level	85.0	50.0
Employment		
Employed	49.3	23.9
Work inside home	46.1	20.6
Health insurance	*	*
None	50.8	14.1
Medicare, Medicaid	33.3	21.4
Private	54.3	34.0
Marital status		
Married	49.6	22.9
Not married	41.0	18.5

[a]Totals may vary due to missing data.
*$P < .05$.

TABLE 3—Prevalence (%) of Pap Smear Screening within Past 2 Years among Mexican-American Women Age 40+ Years, by Acculturation Level, El Paso, Texas, 1991

	Acculturation Level			
	Low 1	2	3	High 4
English proficiency*	42.6	45.0	41.9	57.0
English vs Spanish usage*	43.1	46.3	56.5	57.1
Value placed on cultural origin	48.2	51.4	44.4	37.0
Attitude toward traditional family*	52.2	44.6	46.1	. . .
Interaction with mainstream society*	45.7	46.5	50.0	. . .

[a]Test for trend was restricted to levels 1–3 because of sample sizes less than 5.
*Test for trend $P < .05$.

erty level, <200% federal poverty level, >200% federal poverty level), employment (no, yes), and health insurance (none, medicaid/medicare, private). To eliminate confounding among acculturation dimensions, the final model for each dimension of acculturation included as covariates all other acculturation dimensions (levels 1–4) and sociodemographic factors.

Results

Eighty percent of the women in the study were interviewed in Spanish and 61% were born in Mexico. Only 18% of the women had graduated from high school, 84% had incomes below the federal poverty level, and 45% had no health insurance. Seventeen percent had never had a Pap smear and 65% had never had a mammogram. The median age was 57 years. Sixty two percent of the women were married, and only 17% were employed outside the home. As shown in Table 1, the distribution of women in the study along the five dimensions of acculturation varied widely. Based on the percentages of women in the highest acculturation level, women tended to be more acculturated on English proficiency (24.4%) and personal value placed on preserving Mexican cultural origin (11.1%). The El Paso women in the study were far less acculturated on English language use (3.4%) than on English proficiency (24.4%), which is perhaps not surprising in a border community. Fewer than 1% of the women were highly acculturated with regard to attitude toward traditional family structure and sex-role organization and interaction with members of mainstream society.

Table 2 shows differences in prevalence of Pap smear and mammogram screening in the previous 2 years by various sociodemographic characteristics. The percentage of women who had a recent screening Pap smear declined with each 10-year age group ($P < .05$). Mammogram screening levels for all age groups were consistently low. Women with lower education and income levels were screened less often for cervical and breast cancer. Women on Medicare or Medicaid were least likely to be screened by a Pap smear, even compared with women who have no insurance (a group who may have access through public health clinic gynecologic services). That was not true for mammography; women with no health insurance were least likely to be screened.

The prevalence of Pap smear and mammogram screening by acculturation levels is shown in Tables 3 and 4. The prevalence of Pap smear and mammogram screening increased with each successively higher level of acculturation on proficiency in English and English vs Spanish usage. The other three acculturation dimensions failed to have significant linear effects on Pap smear and mammogram screening. On one dimension, attitude toward traditional family structure and sex-role organization, women did not show the expected pattern of increased screening with greater acculturation. Women least acculturated with regard to traditional family ties had slightly higher rates of Pap smear and mammogram screening.

Because each dimension of acculturation was associated with one or more

socioeconomic characteristics, the independent effect of each acculturation dimension after adjusting for sociodemographic factors was assessed by using multiple logistic modeling. Tables 5 and 6 show the unadjusted and adjusted odds ratios (and their 95% confidence intervals) of screening for each level gain in acculturation. The association of English proficiency and English vs Spanish language usage with Pap smear and mammogram screening disappeared or was reduced when adjusted for age, income, insurance, and education. With all acculturation dimensions and sociodemographic factors in the model, only age remained a statistically significant factor and was inversely related to Pap smear screening ($P = .03$). The strongest independent factor affecting mammogram screening was health insurance coverage ($p = .001$). After adjusting for sociodemographic factors and the other acculturation dimensions, a woman's attitude toward traditional family structure had an independent effect on, and was inversely related to, mammogram screening. Women least acculturated with regard to traditional Mexican family attitudes (importance of extended family and male primacy) were more likely to have been screened by a mammogram.

TABLE 4—Prevalence (%) of Mammogram Screening within 2 Years among Mexican-American Women Age 40+ Years, by Acculturation Level, El Paso, Texas, 1991

	Acculturation Level			
	Low 1	2	3	High 4
English proficiency*	17.3	21.1	20.0	30.4
English vs Spanish usage*	17.5	20.0	38.1	33.3
Value placed on culture origin	15.9	21.6	22.5	28.9
Attitude toward traditional family[a]	26.7	20.4	18.7	. . .
Interaction with mainstream society[a]	17.0	28.6	23.5	. . .

[a]Test for trend was restricted to levels 1–3 because of sample sizes less than 5.
*Test for trend $P < .05$.

TABLE 5—Odds Ratios (95% Confidence Intervals) of Having a Screening Pap Smear for Each Level Gain in Acculturation

	Univariate	Adjusted for Sociodemographic Factors[a]	Adjusted for Sociodemographic[a] and other Acculturation Factors
English proficiency	1.17 (1.00, 1.37)	0.98 (0.79, 1.22)	0.93 (0.70, 1.25)
English vs Spanish usage	1.29 (1.03, 1.61)	1.05 (0.78, 1.41)	1.10 (0.74, 1.65)
Value placed on cultural origin	0.87 (0.71, 1.06)	0.89 (0.72, 1.10)	0.95 (0.76, 1.19)
Attitude toward traditional family	0.90 (0.70, 1.17)	0.81 (0.61, 1.07)	0.81 (0.61, 1.09)
Interaction with mainstream society	1.08 (0.82, 1.41)	1.06 (0.79, 1.43)	1.05 (0.76, 1.45)

[a]Sociodemographic factors were age, education, income, and health insurance.

TABLE 6—Odds Ratios (95% Confidence Intervals) of Having a Screening Mammogram for Each Level Gain in Acculturation

	Univariate	Adjusted for Sociodemographic Factors[a]	Adjusted for Sociodemographic[a] and other Acculturation Factors
English proficiency	1.25 (1.04, 1.51)	1.07 (0.83, 1.39)	0.93 (0.65, 1.32)
English vs Spanish usage	1.53 (1.19, 1.96)	1.32 (0.94, 1.86)	1.28 (0.80, 2.05)
Value placed on cultural origin	1.25 (0.98, 1.60)	1.21 (0.94, 1.57)	1.26 (0.96, 1.66)
Attitude toward traditional family	0.80 (0.58, 1.10)	0.75 (0.54, 1.06)	0.69 (0.48, 0.99)
Interaction with mainstream society	1.37 (0.99, 1.89)	1.20 (0.85, 1.69)	1.14 (0.78, 1.68)

[a]Sociodemographic factors were age, education, income, and health insurance.

Discussion

Although the importance of assimilation in explaining the underutilization of Pap smears and mammograms in Mexican-American women is generally acknowledged, no study has demonstrated an independent effect of Mexican cultural attitudes or values on cancer screening behavior. This study confirms the often-observed finding that language use is an important factor in identifying groups of Hispanic women in need of cancer screening services.[2,4,6,11] In this study the effects of English language proficiency and use disappeared when adjusted for income, education, and health insurance coverage, suggesting a strong correlation with these sociodemographic factors. Therefore, language preference, proficiency, and use measure access to services and are proxies more of socioeconomic status than of cultural attitudes and values.

An important finding of this study is the independent positive effect of Mexican family attitudes and values on mammogram screening. The few studies of Hispanic women and screening behavior have not evaluated cultural values or family attitudes.[2,4,6,11] In a study of elderly Hispanic women living in Los Angeles, recency of mammogram was more frequent in those who had high homeland contact compared with those who had low homeland contact after adjusting for education and age.[6] An interpretation that those women with more homeland contact had stronger family ties provides

more evidence for the positive effect of family on screening. Why Mexican-American women with stronger traditional family attitudes would avail themselves of mammogram screening more frequently may be explained by the support provided by extended family ties either through emotional well-being or shared economic resources. Pap smear screening was not as strongly associated with family attitudes as mammography was. That may be due to the more prevalent use of Pap smears compared with the relative newness of the mammogram in this population.

The same positive effect of family on mammogram screening might also be evident in non-Hispanic women if family attitudes were measured. However, the acculturation scale of family attitudes used in this study was chosen because of its ability to discriminate non-Hispanic White attitudes from Mexican-American attitudes.[8] Family and familism are core characteristics of Hispanic culture.[12] Individuals can acculturate and assimilate remarkably well on language proficiency and use, come to value US traditions more than Mexican traditions, and still retain strong family ties and a high level of perceived family support.[8,12]

The significant effect of family attitudes was demonstrated in a border population not as highly acculturated as the population in San Antonio, Texas, where the acculturation scale was developed.[8] Mexican-American women living in El Paso may not be representative of other US Hispanic populations; in El Paso the Hispanic culture is as dominant as the non-Hispanic culture. It is important to investigate whether this familial effect is observed in Hispanic populations with greater variation on this dimension.

Hispanics are the fastest-growing population in the United States, and Mexican Americans are by far the largest ethnic group. Studies to improve cancer screening in this population consistently equate cultural differences with barriers to screening. This study identifies the positive influence on mammogram screening behavior of one important cultural attribute. Designing effective cancer control interventions for Mexican-American women not only requires taking advantage of the role of family in this group but also should encourage retainment of these cultural values. Future studies should examine this association in other Hispanic populations and, more importantly, examine how the role of familism can be used to change health behavior. □

Acknowledgments

This study was funded by a Public Health Approaches to Breast and Cervical Cancer Screening grant (CA52977) from the National Cancer Institute, Division of Cancer Prevention and Control.

References

1. *Cancer Statistics Review 1973–1986 [Including a Report on the Status of Cancer Control, May 1989]*. Washington, DC: National Cancer Institute; 1989. II51–53. NIH publication 89-2789.
2. Harlan LC, Bernstein AB, Kessler LG. Cervical cancer screening: who is not screened and why? *Am J Public Health*. 1991;81:885–890.
3. Fox SA, Stein JA. The effect of physician-patient communication on mammography utilization by different ethnic groups. *Med Care*. 1991;29:1065–1082.
4. Solis JM, Marks G, Garcia M, Shelton D. Acculturation, access to care, and use of preventive services by Hispanics: findings from HHANES 1982–84. *Am J Public Health*. 1990;80(suppl):11–19.
5. Trevino FM, Moyer ME, Valdez RB, Stroup-Benham CA. Health insurance coverage and utilization of health services by Mexican Americans, Mainland Puerto Ricans, and Cuban Americans. *JAMA*. 1991;265:233–237.
6. Marks G, Solis J, Richardson JL, Collins LM, Birba L, Hisserich JC. Health behavior of elderly Hispanic women: does cultural assimilation make a difference? *Am J Public Health*. 1987;77:1315–1319.
7. Negy C, Woods DJ. The importance of acculturation in understanding research with Hispanic-Americans. *Hispanic J Behav Sci*. 1992;14:224–247.
8. Hazuda HP, Stern MP, Haffner SM. Acculturation and assimilation among Mexican Americans: scales and population-based data. *Soc Sci Q*. 1988;69:687–706.
9. Hazuda H. Items and scoring procedures for the San Antonio Heart Study acculturation and assimilation scales. Unpublished monograph; 1988; San Antonio, Tex.
10. Statistics and Epidemiology Research Corp. EGRET. Seattle, Wash: Statistics and Epidemiology Research Corp; 1990.
11. Stein JA, Fox SA. Language preference as an indicator of mammography use among Hispanic women. *J Natl Cancer Inst*. 1990; 82:1715–1716.
12. Sabogal F, Marin G, Otero-Sabogal R, Marin BV, Perez-Stable EJ. Hispanic familism and acculturation: what changes and what doesn't? *Hispanic J Behav Sci*. 1987;9:397–412.

Risk Behavior and Correlates of Risk for HIV Infection in the Dallas County Household HIV Survey

Geraldine M. McQuillan, PhD, Trena M. Ezzati-Rice, MS, Arlene B. Siller, BS, Wendy Visscher, PhD, and Peter Hurley, BS

ABSTRACT

Objectives. The Dallas County study of a proposed national household seroprevalence survey was designed to assess the feasibility of conducting a national survey and to estimate the prevalence of human immunodeficiency virus (HIV) and hepatitis B virus infection for Dallas County. Risk behavior data were collected and correlated with HIV infection.

Methods. Participants in this survey represented a probability sample of the county. A self-administered questionnaire on demographic characteristics and HIV risk behavior was completed and a blood sample was obtained.

Results. Of the 1724 adults eligible for the survey, 1446 completed the questionnaire and 1374 provided a blood sample. The prevalence estimates were 0.4% for HIV and 7.3% for hepatitis B virus. A strong relationship was observed between HIV and hepatitis B status and risk behavior.

Conclusions. In this study population, receptive anal intercourse and increasing numbers of male partners had the strongest correlation with the prevalence of HIV and hepatitis B virus infection in men. The high level of risk reporting for individuals positive for HIV or hepatitis B suggests that survey participants who engage in risk behaviors were willing to report those behaviors. (*Am J Public Health.* 1994;84:747–753)

Introduction

Estimating the magnitude and direction of the spread of human immunodeficiency virus (HIV) in the general population is of considerable public health interest. However, as reported by Massey et al.,[1] there are numerous technical and methodological issues involved in obtaining a direct estimate of the rate of HIV infection in the general population. Estimates of the size of HIV risk groups, their distribution in the population, and their relationship to HIV infection can be useful in modeling the HIV/acquired immunodeficiency syndrome (AIDS) epidemic; they can also be useful for effective planning and allocation of resources for educational and prevention programs to reduce the numbers of persons engaging in HIV risk behaviors.

The Dallas County Household HIV Survey, sponsored by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), was conducted in fall 1989 by the Research Triangle Institute. In addition to assessing the technical feasibility of a national household HIV survey, the Dallas study, which used a probability sample of the county, was designed to estimate the extent of HIV infection among adults 18 to 54 years of age. Further, the survey was designed to provide information on the distribution of HIV risk behaviors and correlates of risk among subgroups of the population for use in planning local HIV prevention and treatment programs.

In addition to estimates of self-reported risk behaviors and correlates of risk, this paper examines the relationship of HIV and hepatitis B virus prevalence estimates for selected subgroups of the population defined by demographic characteristics and risk behavior. Antibody to hepatitis B core antigen, a serologic marker for past or present infection with hepatitis B virus, was included in this survey as a surrogate measure to help evaluate the level of risk behaviors reported. These data, based on a representative sample, provide information on HIV infection and risk behaviors for noninstitutionalized adults in Dallas County.

Methods

Sample Design and Selection

The target population was residents of households of Dallas County, which was selected for the study because it had a sufficiently high AIDS case rate and its population size could yield for analyses an adequate number of HIV-positive persons who were 18 to 54 years of age and not on active military duty. The survey used a stratified area probability design, with all households in the county having a chance of selection. The sampling frame was based on the count of housing units from the 1980 census, updated to be representative of Dallas County in 1989. The area sampling frame was geographically stratified by race/ethnicity, sex, and marital status and by expected level of HIV risk. The construction of three HIV

Geraldine M. McQuillan, Trena M. Ezzati-Rice, Arlene B. Siller, and Peter Hurley are with the National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Md. Wendy Visscher is with the Research Triangle Institute, Research Triangle Park, NC.

Requests for reprints should be sent to Geraldine M. McQuillan, PhD, National Center for Health Statistics, 6525 Belcrest Rd, Room 1070, Hyattsville, MD 20782.

This paper was accepted July 19, 1993.

Editor's Note. See related annotation by Francis (p 720) in this issue.

TABLE 1—Weighted Response Rates for Sample Persons Completing the Questionnaire and Giving Blood, Dallas County, Texas, 1989

	% of Population	Questionnaire[a]		Blood and Questionnaire[a]	
		%	P (χ^2)	%	P (χ^2)
Risk stratum			.11		.94
High	7.7	83.4		83.2	
Medium	21.7	89.3		84.6	
Low	70.6	90.5		83.9	
Age, y			.02		.41
18–24	19.9	96.2		87.3	
25–34	38.0	91.7		85.1	
35–54	42.3	85.1		81.5	
Sex[b]			.05		.37
Male	48.2	87.5		82.1	
Female	51.8	91.9		85.8	
Race and sex			.90		.85
Black total	17.2	94.2		88.9	
Male	7.2	95.1		94.4	
Female	10.0	93.6		84.8	
Hispanic total	14.8	91.6	.90	80.2	.40
Male	7.7	91.1		82.2	
Female	7.1	92.2		78.1	
Other total	68.1	88.3	.95	83.6	.50
Male	33.4	85.3		79.3	
Female	34.7	91.4		87.6	
Marital status[b]			.90		.01
Married	59.8	89.9		80.1	
Not married	40.2	89.5		89.7	
Total	100	89.7		84.0	

[a]The response rates in this table do not include the 2% screening nonresponse.
[b]Sex and marital status were missing for one sample person each among persons giving blood; sex was missing for one sample person; marital status was missing for two sample persons among persons completing the questionnaire.

risk strata (high, medium, or low) was based on summary public health statistics of HIV risk indicators, including sexually transmitted disease prevalence and AIDS case distribution.[2] This stratification by expected HIV risk level was designed to ensure that both high- and low-risk persons would be adequately represented in the sample. The sample was balanced within each of the three risk strata so that the proportion of Black, Hispanic, and White households was approximately equal to the overall race/ethnicity distribution in Dallas County. A detailed description of the sample design has been published elsewhere.[2]

Survey teams consisting of an interviewer and a phlebotomist visited each sample household to determine if it contained at least one eligible adult. One person was selected at random from each age-eligible household. After agreeing to participate, sample persons were asked to complete a 10-minute self-administered questionnaire containing basic demographic questions as well as questions about behaviors related to the risk of HIV infection; individuals were asked to report those behaviors practiced since 1978 and within the past 12 months. Respondents either completed the questionnaire in private or had the interviewer read the questionnaire to them while they marked their own answers. The certified phlebotomist then drew a blood sample.

Risk-behavior data and blood sample results were not linked to personal identifiers. Other strict privacy procedures implemented in the survey to ensure a completely anonymous dataset have been previously described.[2] Sample persons who provided both a blood sample and a completed questionnaire in the regular survey received a $50 cash payment. In addition to the regular survey, a special quality assessment study was conducted following the main survey in an attempt to increase the overall survey response rate and to provide data for assessing and adjusting for potential nonresponse bias.[2] Data from both the regular study and the quality assessment study are combined for the analyses in this paper.

Laboratory Analyses

Laboratory analysis of the sample included assays for HIV antibody (enzyme-linked immunosorbent assay, Dupont, Inc) and for the antibody to hepatitis B core antigen (Abbott Laboratories, North Chicago, Ill). Any reactive specimen for HIV antibody was retested and confirmed by Western blot analysis (Dupont, Inc).[3]

Statistical Analyses

Estimates of the prevalence of selected HIV risk behaviors for persons aged 18 to 54 years in Dallas County are based on sampling weights that were adjusted for household screening and sample person nonresponse. To adjust for nonresponse, basic sampling weights were first corrected to compensate for screening and questionnaire nonresponse. A further nonresponse adjustment was made in the weights for the HIV and hepatitis B virus prevalence estimates by imputing HIV and hepatitis B infection probabilities for sample persons who lacked blood test results. This included persons who did not provide a blood sample in the quality assessment study, as well as those whose blood samples were not sufficient for laboratory analysis. The two prevalence estimates were formed as a weighted average of the actual HIV antibody test results and imputed HIV and hepatitis B infection probabilities.[2] Software for survey data analysis (SUDAAN) was used to develop the prevalence estimates and distributions, taking into account the adjusted sampling weights and complex survey design.[4] The chi-square test was used to test for statistical significance at the $P = .05$ level. Percentage of population columns in the tables reflect the distribution of demographic or behavioral characteristics of the survey participants.

Quality of Data

Of the sample person questionnaires (n = 1446), 83% had no blank items or unusable responses. The highest item nonresponse rate (6%) was for the question that asked how often a condom was used during sex with a woman in the past 12 months. A 4% item nonresponse rate was obtained for the same question asked for the time period since 1978. Since more women did not answer these questions than men, these items were probably left blank by women who felt that the items

TABLE 2—Estimates of Reported HIV Risk Behaviors since 1978, by Demographic Characteristics, Dallas County HIV Survey, 1989

Risk Factor	Age, y			Race/Ethnicity			Sex		Marital Status		Education[a]		
	18–24 (n = 276)	25–34 (n = 626)	35–54 (n = 544)	Black (n = 384)	Hispanic (n = 207)	Other (n = 855)	Male (n = 701)	Female (n = 745)	Married (n = 573)	Not Married (n = 873)	<12 y (n = 263)	12+ y (n = 1165)	Total (n = 1446)
Male-to-male sex, %	1.7	11.4	6.9*	2.6	13.5	6.7	7.6	NA	4.6	11.5	9.6	7.1	7.6
Receptive anal intercourse, %	8.3	16.2	11.2*	5.8	13.9	13.9*	4.5	20.4*	11.2	13.8	16.5	11.6	12.6
Intravenous drug use (women and heterosexual men), %	4.0	7.1	1.0*	3.0	1.4	4.7	5.2	2.7	1.2	6.0*	6.2	3.3	3.9
Received money or drugs for sex, %	0.5	4.1	0.7*	4.4	3.8	0.8*	1.8	2.2	1.1	3.2	2.2	1.9	2.0
Gave money or drugs for sex, %	4.1	7.8	5.6	5.5	10.3	5.2*	11.6	0.9*	4.4	8.7	5.8	6.3	6.2
Heterosexual contact													
Sex with intravenous drug abuser, %	12.3	9.2	6.9	10.5	4.3	9.7*	8.2	9.2	6.2	12.5*	12.0	8.0	8.8
Sex with bisexual men (women only), %	4.3	2.9	1.6	1.8	3.2	2.4	NA	5.2	2.0	3.4	1.6	2.9	2.6
10+ partners, %	14.8	18.8	7.6*	12.6	9.7	14.5*	19.2	7.6*	9.8	18.1*	8.7	14.4	13.2

[a]Data on education were missing for 18 sample persons.
*Chi-squared test for homogeneity within the demographic categories, $P < .05$.

did not apply to them. The intravenous drug use questions had a nonresponse rate of less than 1%, among the lowest observed. Fifty-seven persons (4%) had at least 18 blank items or did not answer key risk behavior questions; these persons were more likely to be from the high-risk stratum, older (35 to 54 years of age), and Hispanic. Item nonresponse data were treated as missing in the analyses.

Results

Participation Rates

Of the 1724 adults aged 18 to 54 years identified as eligible for the Dallas County Household HIV Survey, close to 90% participated; 1446 completed the questionnaire and 1374 provided a blood sample. Two sample person rates are shown in Table 1 since the special quality assessment study following the regular survey included two randomly assigned treatment groups. Half the persons were asked to complete the questionnaire only and the other half were asked to complete the questionnaire and provide a blood sample, the same procedure used in the regular survey. The weighted response rate for persons completing the questionnaire in both the regular survey and the quality assessment study was 90%, while the response rate for individuals completing a questionnaire and providing a blood sample was 84%. These response rates equal or exceed those attained in other national health surveys.[5] There were no differences in the response rates for questionnaire or blood and questionnaire among the three HIV risk strata. The lowest blood sample and questionnaire response rates were for men who reported that their ethnic group was neither Black nor Hispanic and who therefore were considered to be in the White or other ethnic group and for Hispanic women. Married individuals were less likely to provide both a blood sample and a questionnaire, and men and individuals in the oldest age group (35 to 54 years) were less likely to answer the questionnaire. If marital status is stratified by race and sex, married men and married Hispanics had significantly higher nonresponse to both the blood sample and the questionnaire.

Table 2 presents the self-reported risk behaviors by demographic categories. Overall, an estimated 8% of men aged 18 to 54 years in Dallas County reported that they engaged in male-to-male sex or bisexual activity since 1978. Receptive anal intercourse was reported by almost 5% of the men and 20% of the women. The 25- to 35-year age group, men, and not married participants were more likely to report HIV risk behaviors.

The distribution of selected health conditions and behaviors that may be correlated with HIV infection are presented in Table 3. These data, in addition to providing estimates of selected correlates of risk, also provide some indication of whether persons at risk for HIV infection participated in the survey. The rate of reported gonorrhea in the high-risk stratum was 1.7 times that reported by persons in the medium- or low-risk stratum. There was no difference between the medium- and low-risk strata for most of the correlates of risk except for reporting ever having had an HIV antibody test. Thirteen percent of the men and 8% of the women reported having been told by a doctor that they had gonorrhea. A significant racial difference was observed: 22% of Black participants reported a diagnosis of gonorrhea, compared with 9% of Hispanics and 8% of Whites. When data for gonorrhea, syphilis, and genital herpes were combined (any sexually transmitted disease), the same racial/ethnic pattern was observed as well as the higher rate in the high-risk stratum. Approximately one out of seven persons reported having previously been tested for HIV. The level of testing for persons in the high- and medium-risk strata was twice that for persons in the low-risk stratum. Twelve percent of survey participants reported "usually or every time" using a condom during sex with a man. These data reflect condom use for heterosexual women and

TABLE 3—Correlates of Risk for HIV Infection, Dallas County, Texas, 1989, by Demographic Characteristics

	Ever Had Gonorrhea		Ever Had Any Sexually Transmitted Disease[a]		Ever Tested for HIV Antibody		Condom Use during Sex with Male[b]	
	%	$P(\chi^2)$	%	$P(\chi^2)$	%	$P(\chi^2)$	%	$P(\chi^2)$
Percentage of total population	10.3		13.7		14.9		12.4	
Risk stratum		.01		<.001		<.001		.53
High	16.5		19.8		25.9		12.4	
Medium	9.8		14.7		22.7		14.9	
Low	9.7		12.7		11.3		11.6	
Age, y		.12		.54		<.001		<.001
18–24	5.3		11.0		12.5		28.1	
25–34	9.9		14.2		18.3		13.6	
35–54	12.9		14.5		12.8		5.7	
Race/ethnicity		<.001		<.001		.40		.01
Black	21.8		26.7		12.8		21.4	
Hispanic	9.4		12.0		16.0		4.1	
Other	7.6		10.9		15.1		11.9	
Sex		.05		.48		.90		.23
Male	12.9		15.5		15.0		7.5	
Female	7.7		12.0		14.8		12.8	
Marital status		.05		.03		.88		.02
Married	8.4		10.5		14.7		9.0	
Not married	12.8		16.2		15.1		17.6	
Education		.66		.91		.60		.16
<12 y	11.3		14.7		13.8		7.7	
12+ y	10.0		13.5		15.2		13.5	

[a]Gonorrhea, syphilis, or genital herpes.
[b]Women and men reporting male-to-male sex and "usually or every time" using a condom.

TABLE 4—HIV and Hepatitis B Virus Prevalence Estimates by Demographic Characteristics, Dallas County, Texas, 1989

		HIV Positive		Hepatitis B Virus Positive	
	n	%	SE	%	SE
Risk stratum					
High	401	2.2	0.9	11.3	1.7
Medium	578	0.8	0.4	8.5	1.1
Low	467	0.1	0.1	6.4	1.6
Age, y					
18–24	276	0.4	0.3	1.5	0.5
25–34	626	0.7	0.3	7.4	1.3
35–54	544	0.2	0.1	9.8	2.4
Race/ethnicity					
Black	384	0.5	0.4	10.0	1.5
Hispanic	207	0.4	0.4	2.3	1.0
Other	855	0.4	0.1	8.0	1.7
Sex					
Male	701	0.7	0.2	7.7	1.1
Female	745	0.1	0.1	6.9	1.7
Marital status					
Married	573	0	0	6.9	1.8
Not married	873	1.0	0.31	7.8	1.1
Total	1446	0.4	0.1	7.3	1.2

for men reporting male-to-male sex. There was no difference in the reporting of this behavior across risk strata.

Relationship of HIV Status with Self-Reported Risk Behaviors

The overall weighted prevalence of HIV was 0.4% and of hepatitis B virus was 7.3% (Table 4). Both virus estimates in the high-risk stratum were significantly higher than those in the low-risk stratum. Regarding HIV prevalence alone, no differences were seen between ethnic and racial subgroups. Patterns seen across the other demographic variables are consistent with known patterns of HIV infection.[6] HIV prevalence was significantly higher among men and among individuals who reported that they were not married. Finally, the highest HIV prevalence among the three age groups was for the 25- to 34-year group.

The prevalence of hepatitis B virus was also consistent with known patterns of infection.[7] There was no difference in estimates between men and women, and no association with marital status. Prevalence was significantly higher among persons in the two older age groups than among persons aged 18 to 24, while Hispanics had significantly lower hepatitis B virus estimates than other racial and ethnic subgroups.

A strong relationship was observed between the HIV status of the participants and the self-reported risk behavior data. Of the 15 sample persons who were HIV positive, all 13 men and 1 woman reported at least one behavior known to increase the risk of HIV infection. The second HIV-positive woman reported only five to nine heterosexual partners since 1978 as a potential risk factor. All 15 HIV-positive individuals were in the high- and medium-risk strata. Overall, 73% of the 120 individuals who were positive for hepatitis B virus reported one or more risk behaviors associated with HIV infection or some correlate of risk.

HIV prevalence estimates by key risk behaviors and correlates of risk are presented in Table 5. Among men who reported male-to-male sex, there was a positive relationship between the number of sex partners and the prevalence of both HIV and hepatitis B virus infection. Among men who reported 10 or more male partners, the estimated prevalence of HIV infection was almost 23% and the prevalence of hepatitis B virus was even higher (62%). Intravenous drug use was associated with a significant increase of hepatitis B virus but not of HIV in this

population. Although only 4.5% of the men reported engaging in receptive anal intercourse, 14% of these individuals were infected with HIV. Twenty percent of the women reported engaging in receptive anal intercourse, with Black women reporting a significantly lower rate of this behavior. Although this behavior was not associated with an increase in HIV infection in these women, it was associated with an increase in hepatitis B virus infection. Women reporting this behavior had twice the prevalence and men had four times the prevalence of hepatitis B virus compared with the total population.

Table 6 presents HIV and hepatitis B prevalence estimates by correlates of risk. Condom use appears to be a good predictor of both infections in men; no significant association was seen for women. An equal percentage of men and women (15%) reported ever having been tested for HIV antibody. Men who reported prior HIV testing had a higher prevalence of HIV infection than all men. Serologic evidence of hepatitis B virus infection was strongly correlated with elevated HIV infection in men.

TABLE 5—HIV and Hepatitis B Virus Prevalence Estimates by Key Risk Factors since 1978, Dallas County, Texas, 1989

Risk Factor	% of Population	SE	HIV Infection %	HIV Infection SE	Hepatitis B Virus Infection %	Hepatitis B Virus Infection SE
Total population			0.4	0.1	7.2	1.2
Males	49.3	2.5	0.7	0.2	7.7	1.1
Females	50.7	2.5	0.1	0.1	6.9	1.7
Male-to-male sex: number of male partners						
1–4	5.4	1.8	1.4	1.0	11.9	3.2
5–9	11.3	7.9	11.1	1.0	28.6	4.2
10+	19.2	4.8	22.8	6.6	61.8	10.5
Intravenous drug use, no male-to-male sex	3.9	0.9	1.0	0.6	28.7	8.1
Receptive anal intercourse						
Total population	12.6	1.6	2.5	0.8	17.6	6.2
Males	4.5	1.5	14.0	3.1	33.4	5.6
Black	1.1	0.8	0.0	0.0	0.0	0.0
Hispanic	10.1	6.5	8.5	2.6	23.1	4.0
Other	3.6	0.9	19.1	4.9	43.4	7.7
Females	20.4	2.2	0.0	0.0	14.2	7.6
Black	9.0	2.5	0.0	0.0	8.6	3.7
Hispanic	16.1	3.5	0.0	0.0	0.0	0.0
Other	24.3	3.1	0.0	0.0	17.6	9.7
Heterosexual contact						
Sex with intravenous drug user	8.8	1.3	0.3	0.2	12.0	4.1
Number of female partners for men						
1–4	49.7	3.9	0.3	0.2	6.5	1.3
5–9	21.0	2.5	0.3	0.2	5.8	2.9
10+	19.2	2.1	0.2	0.1	10.4	4.1
Number of male partners for women						
1–4	70.3	2.7	0.2	0.15	6.7	2.3
5–9	13.8	7.9	0.2	0.14	9.3	2.2
10+	7.6	1.1	0.0	0.0	7.6	3.9

Discussion

A survey based on a probability sample of US households had been suggested as a method to determine the number of persons infected with HIV.[8] To test the feasibility of such a survey, the NCHS conducted pilot studies at two sites through a contract with the Research Triangle Institute. This report summarizes the second field test, which was conducted in Dallas County, Texas, from September through December 1989. The response rates achieved in the Dallas survey equal or exceed those in other national health surveys.[5]

The level of risk reporting among sample persons positive for both viruses indicate that persons who engage in risk behaviors for HIV and hepatitis B virus infection will report these behaviors in a self-administered anonymous questionnaire. However, the extent of underreporting of these risk behaviors cannot be readily determined because there are no comparable studies for Dallas County, nor are there established norms in the general population. Nevertheless, the risk behaviors reported do provide information for targeting public health messages in Dallas County to prevent HIV transmission.

In particular, the one risk behavior with an unusually high level of reporting, receptive anal intercourse, should be more explicitly addressed in educational messages to the general public. Although 12.6% of the sample persons selected for the survey reported engaging in receptive anal intercourse, this risk behavior was much higher in women (20.4%). This is a higher percentage than is usually reported for this behavior in the US population.[9] Stratifying by race indicated that women who were in the non-Hispanic White/other ethnic group were one-and-a-half times more likely than Hispanic women and almost three times more likely than Black women to report this behavior. This risk behavior, although not associated with HIV infection in women at the time of the survey, does place these women at increased risk of infection as the transmission of the virus moves to the heterosexual population. This potential risk is demonstrated by the twofold higher hepatitis B virus prevalence in women in the non-Hispanic White/other ethnic group who reported this behavior compared with the total population. Thus, these women in Dallas County need to be especially targeted for messages on the risk of unprotected receptive anal sex.

Overall, 12.4% of the sample (men and women) who reported having sex with a man used condoms. This estimate is similar to other population-based surveys in the United States regardless of reporting of AIDS-related risk factors.[10]

Ten percent of the population reported having been diagnosed with gonorrhea. This reported rate in the Dallas County population is higher than that reported to the CDC.[11] In 1990, the total incidence rate of gonorrhea in the United States was 266 6 per 100 000 population (0.27%). If one accumulated reported cases between 1978 and 1989 (the time frame asked in the Dallas survey), the reported rate in the United States would

TABLE 6—HIV and Hepatitis B Virus Prevalence Estimates by Correlates of Risk,[a] Dallas County, Texas, 1989

Risk Factor	% of Population	SE	HIV Infection %	HIV Infection SE	Hepatitis B Virus Infection %	Hepatitis B Virus Infection SE
Total population			0.4	0.1	7.3	1.2
Males	49.3	2.5	0.7	0.2	7.7	1.1
Females	50.7	2.5	0.1	0.1	6.9	1.7
Condom use during sex with man[b]						
Males	7.5	3.4	16.4	4.0	40.8	8.4
Females	12.8	1.7	0.1	0.1	9.4	3.5
Total	12.4	1.7	1.2	0.4	11.6	3.3
History of sexually transmitted disease[c]						
Males	14.9	2.3	1.3	0.7	12.4	3.0
Females	10.9	1.5	1.0	1.0	11.2	3.9
Total	12.9	1.1	1.2	0.6	11.9	2.4
Ever tested for HIV						
Males	15.0	1.9	3.0	1.0	10.5	2.6
Females	14.3	2.1	0.0	0.0	3.5	1.1
Total	14.9	1.3	1.5	0.5	7.0	1.2
Hepatitis B infection						
Males	7.7	1.1	6.3	2.6	...	...
Females	6.9	1.2	0.0	0.0	...	...
Total	7.3	1.1	2.8	1.2	...	...

[a]Risk reporting is from 1978 to present.
[b]Women and men reporting male-to-male sex and "usually or every time" using a condom.
[c]Gonorrhea, syphilis, or genital herpes.

be 5.0%.[12] However, it is well accepted that there is considerable underreporting of gonorrhea cases to the CDC, which these data may further substantiate. A recent overview on gonococcal infections in the United States estimates that the reported cases represent half the actual infections.[13] In addition, the male-to-female ratio in Dallas (1.7:1) is similar to that for the total United States (1.5:1).[11] If these assumptions of national reporting are correct, the 10% prevalence of self-reported gonorrhea is in the range expected.

The overall prevalence of HIV infection among the Dallas County household population was 0.4%. The prevalence of infection was significantly higher in men and in unmarried participants. There was also a strong relationship between self-reported risk behavior and HIV and hepatitis B virus prevalence; however, there was no difference between the medium- and low-risk strata for most of the correlates of risk except for reporting ever having had an HIV antibody test. The risk strata were developed using summary statistics obtained from the Dallas County Health Department surveillance programs. Population segments were classified as high, medium, or low, based on their ranking from high to low on either a rate or relative number of reportable conditions (AIDS, syphilis, hepatitis B, etc.). These summary health statistics were useful in separating the highest risk areas for sampling but less useful for differentiating moderate- from low-risk areas.

The patterns of prevalence of HIV infection across demographic variables are consistent with known patterns of HIV infection.[8,14-16] HIV prevalence was significantly higher among men than among women, as well as among individuals who reported that they were not married compared with those who reported that they were married. Finally, the highest prevalence among the three age groups was in the 25- to 34-year group. The patterns of prevalence by key risk factors were also consistent with known HIV risk behaviors.[6] Increased numbers of male-to-male sexual partners and the practice of receptive anal intercourse among men increased the prevalence of infection. Condom use during sex with a man was associated with an increase in HIV and hepatitis B virus prevalence in men, perhaps because men who knew they were infected or at risk for infection used condoms more often to protect their partners. The prevalence of hepatitis B across demographic variables was also consistent with known patterns.[7]

The strong correlation of risk factors with HIV and hepatitis B infection support the assumption that sample people are willing to report risk behavior. Although analysis of the epidemiology of these infections does not provide evidence that all high-risk individuals participated in the Dallas survey, it does suggest that those who participate in high-risk behavior were included in the survey and were willing to report this behavior. Further, the strong correlation between risk behavior reporting and HIV status among blood and questionnaire respondents provided the basis for predicting the HIV status of persons who did not provide a blood sample.

These results demonstrate that it is technically feasible to conduct a national household HIV survey, as demonstrated by the success of using phlebotomist-interviewer teams to collect blood samples and self-reported risk behavior data. However, despite the success in fielding the study and evidence that individuals who engage in high-risk behavior would participate in the survey, the inability to directly estimate nonresponse bias in the HIV prevalence estimate became an overriding concern. Because of this concern, there was no clear consensus on the scientific merit of a national survey when the CDC asked a number of outside public health, epidemiological, and survey experts to review and comment on the Dallas County HIV survey plan and findings. Therefore, the director of the NCHS recommended to the director of the CDC that a national survey not be done at the present time.[17]

In spite of this, however, the risk behavior data reported in this survey from a representative sample of Dallas County provides both a basis for community education/prevention programs to prevent the transmission of HIV infection and a useful comparison for similar surveys conducted in other counties in the future. □

References

1. Massey JT, Ezzati TM, Folsom R. Statistical issues in measuring the prevalence of HIV infection in a household survey. In: *1990 Proceedings of the Section of Survey*

Research Methods of the American Statistical Association; July 1991:160–169.
2. *National Household Seroprevalence Survey Feasibility Study Final Report.* Research Triangle Park, NC: Research Triangle Institute; April 30, 1990. Research Triangle Report RTI/4190-01/01F.
3. Centers for Disease Control. Interpretation and use of the Western blot assay for serodiagnosis of human immunodeficiency virus type 1 infections. *MMWR.* 1989;38(S-7):1–7.
4. *SUDAAN.* Research Triangle Park, NC: Research Triangle Institute; 1991.
5. *Plan and Operation of the Second National Health and Nutrition Examination Survey 1976–1980.* Hyattsville, Md: National Center for Health Statistics; 1981:1–144. Programs and collection procedures, series 1, no. 15.
6. Centers for Disease Control. The HIV/AIDS epidemic: the first 10 years. *MMWR.* 1991;40:357–369.
7. McQuillan GM, Townsend TR, Fields HA, et al. Seroepidemiology of hepatitis B virus infection in the United States, 1976–1980. *Am J Med.* 1989;87(suppl 3A):5–10.
8. Turner CF, Fay RE. Monitoring the spread of HIV infection. In: Turner CF, Miller HG, Moses LE, eds. *AIDS Sexual Behavior and Intravenous Drug Use.* Washington, DC: National Academy Press; 1989:447–470.
9. Voeller B. AIDS and heterosexual anal intercourse. *Arch Sex Behav.* 1991;20: 233–277.
10. Catania JA, Coates TJ, Stall R, et al. Prevalence of AIDS-related risk factors and condom use in the United States. *Science.* 1992;258:1101–1106.
11. Centers for Disease Control. Summary of notifiable diseases in the US. 1990. *MMWR.* 1990;39:22–24.
12. Centers for Disease Control. Progress toward achieving the 1990 objectives for the nation for sexually transmitted diseases. *MMWR.* 1990;39:53–57.
13. Dalabetta G, Hooks EW. Gonococcal infections. In: Handsfield HH, ed. *Infectious Disease Clinics of North America:* [Vol. 1, no. 1] *Sexually Transmitted Diseases.* Philadelphia, Pa: WB Saunders Co; 1987; 25–54.
14. Rogers SM, Turner CF. Male-male sexual contact in the USA: findings from five sample surveys, 1970–1990. *J Sex Res.* 1991;28:491–519.
15. Fay RE, Turner CF, Klassen AD, Gagnon JH. Prevalence and patterns of same-gender sexual contact among men. *Science.* 1989;243:338–348.
16. Centers for Disease Control. Numbers of sex partners and potential risk of sexual exposure to human immunodeficiency virus. *MMWR.* 1988;37:565–568.
17. Centers for Disease Control. Pilot study of a household survey to determine HIV seroprevalence. *MMWR.* 1991;40:1–5.

Self-Reports of HIV Risk Factors by Patients at a Sexually Transmitted Disease Clinic: Audio vs Written Questionnaires

Bradley O. Boekeloo, PhD, ScM, Laura Schiavo, BA, David L. Rabin, MD, MPH, Richard T. Conlon, MPA, Carol S. Jordan, MPH, BSN, and Diane J. Mundt, PhD

ABSTRACT

Objectives. The purpose of this study was to determine how the method of assessment affects patient report of human immunodeficiency virus (HIV) risks.

Methods. Patients at a sexually transmitted disease clinic randomly received either a written self-administered questionnaire or an audio self-administered questionnaire delivered by cassette player and headset. These questionnaires were followed by face-to-face interviews.

Results. Audio questionnaires had fewer missing responses than written questionnaires. Audio questionnaires also identified more unprotected vaginal intercourse and sexual partners suspected or known to have HIV infection or acquired immunodeficiency syndrome than did written questionnaires. Although both the audio and written questionnaires identified more risks than the face-to-face interviews, the difference in the mean number of reported risks between the audio questionnaires and the face-to-face interviews was greater than that between the written questionnaires and the face-to-face interviews.

Conclusions. Audio questionnaires may obtain more complete data and identify more HIV risk than written questionnaires. Research is warranted about whether audio questionnaires overcome barriers to the completion and accuracy of HIV risk surveys. This study emphasizes the need to elucidate the relative strengths and weaknesses of written questionnaires, audio questionnaires, and face-to-face interviews for HIV risk assessment. (*Am J Public Health.* 1994;84: 754–760)

Introduction

Reliable and valid human immunodeficiency virus (HIV) risk assessment is integral for behavioral research, effective risk reduction counseling as part of clinical practice, and safe blood donation. Measurement error in HIV risk assessment can bias rate estimates of high-risk behaviors, leading to misinterpretation of research findings.[1] In clinical practice such an error can lead to undetected patient risk, resulting in missed opportunities for HIV diagnostic testing and risk reduction counseling or mislabeling of individuals as high or low risk.[2–4] As part of blood donation such an error can weaken efforts to protect the blood supply from HIV contamination.[5]

For most HIV risk factors, which include specific risky sexual and drug behaviors, self-report is typically the only acceptable method of assessment. Both face-to-face interviews[3,6–10] and self-administered questionnaires[1,2,4,11] are common methods of assessing HIV risk factors. There is evidence, however, that face-to-face interviews may have greater measurement error than self-administered questionnaires when obtaining sensitive information like sexual and drug use behaviors,[12] although neither method has been established as superior. Concern that interviews do not accurately obtain sensitive information and that paper-and-pencil self-administered questionnaires (written questionnaires) are difficult for some individuals to read and complete has led to research on computer-assisted self-administered risk assessment methods.[5]

A self-administered questionnaire that is administered by audiocassette player and headset (audio questionnaire) potentially overcomes some of the sources of measurement error associated with face-to-face interview and a typical written questionnaire. Also, an audio questionnaire is less expensive and more adaptable for use by providers in private offices than a computer-assisted assessment. Compared with the written questionnaire, the audio questionnaire might be more likely to hold respondent attention and alleviate reading problems. It might be more confidential than a written questionnaire in a nonprivate administration setting if only response options and no questions appear on the response form. Compared with the face-to-face interview, the audio questionnaire might be perceived as less threatening when behaviors of a sensitive nature are assessed. On the other hand, although the tape can be rewound and the questions reviewed, the pace of the audio questionnaire may be frustrating.

Because the primary problem in HIV risk assessment is underreporting rather than overreporting,[5] an assessment method that identifies more HIV risks than currently accepted methods is needed. Therefore, the major objective of this study was to determine whether an

Bradley O. Boekeloo, Laura Schiavo, and David L. Rabin are with the Department of Community and Family Medicine, Georgetown University School of Medicine, Washington, DC. Richard T. Conlon is with the National Center for Prevention Services, Division of STD/HIV Prevention, Centers for Disease Control and Prevention, Atlanta, Ga. Carol S. Jordan is with the Montgomery County Health Department, Silver Spring, Md. Diane J. Mundt is with the National Academy of Sciences, Institute of Medicine, Washington, DC.

Requests for reprints should be sent to Bradley O. Boekeloo, PhD, ScM, The George Washington University Medical Center, Department of Health Care Services, 2150 Pennsylvania Ave, NW, Washington, DC 20037.

This paper was accepted June 8, 1993.

audio questionnaire identifies more HIV risks than a written questionnaire in a sexually transmitted disease clinic population. The study design further allows preliminary examination of whether the two questionnaire methods identify more risk than follow-up face-to-face interviews.

Methods

The site of this study was a public sexually transmitted disease clinic in a Maryland suburb of Washington, DC. Patients 16 years of age or older who could independently complete a brief clinic registration form that asked for basic demographic information were eligible for the study. Independent completion of the registration form was used as a screen for reading and comprehension ability. Of 1017 clinic patients seen during the 6-month study period, 934 were eligible, 22 were previously enrolled repeat attenders, and 60 were ineligible. Of those who were ineligible, 82% could not independently complete the registration form because they could not understand or read English, 10% were incapable of independently completing the registration form for other reasons such as inebriation, and 8% were under 16 years of age.

The research interviewer always recruited the most recently registered eligible sexually transmitted disease clinic patient for risk assessment, skipping other patients so as not to disrupt clinic flow. Of 328 sexually transmitted disease clinic patients asked to participate in this study, 93% agreed. Patients signed consent forms before study participation and were told that information was being collected for study purposes only. Predetermined code numbers, not names, appeared on questionnaires so that the questionnaires could be matched with the face-to-face interviews. Patients were randomly assigned to receive either the written or audio questionnaire (Figure 1). After the questionnaire was administered to patients, the interview was administered. The questionnaire and interview were conducted privately, one participant at a time.

Questions to assess HIV risk were developed from model questionnaires used in previous research and included patient characteristics, sexually transmitted disease history during the past year, age at onset of sexual behavior, needle sharing ever in the past, and specific sexual, drug, and alcohol behaviors during the past 3 months. A written questionnaire was developed from the questions (Appendix). The written questionnaire was then read verbatim by a Black sexually transmitted disease counselor to provide an audiocassette for use with a headset and audiocassette player. A response form with only question numbers and response options was developed, and this form along with the audiocassette player and headset composed the audio questionnaire. The written and audio questionnaires were tested with four focus groups of volunteer sexually transmitted disease clinic patients: one group each of Black males, Black females, non-Black males, and non-Black females. Based on results from the focus groups, several wording changes were made, such as changing "sexual intercourse" to "sex" when vaginal, oral, and anal intercourse were being discussed because for many participants, "sex" connoted all intercourse and "sexual intercourse" connoted only vaginal intercourse. Short explanatory statements were added preceding questions on prostitution, alcohol use, drug use, and needle sharing because focus group participants thought that unless the relevance of these questions was explained to them, they would hesitate to answer the questions. Also, music was deleted from the audiocassette and a local Black female radio celebrity was chosen as a reader for the final audiocassette in response to preferences of the focus group participants. Revised questionnaires were then pilot-tested with sexually transmitted disease clinic patients before initiation of the study.

The survey questions were written at a seventh-grade reading level as confirmed by the Fry index,[13] whereas the survey instructions tested at at least a high school reading level. Scaled responses and skip patterns added to the difficulty of some questions. A 30-year-old Black male, an experienced counselor in the sexually transmitted disease clinic, administered both the questionnaires and the interviews. Although he was not blinded to the study group of the patient, the interviewer did not review completed questionnaires and was never part of discussions about research questions and hypotheses. Using a blank copy of the written questionnaire as a guide and response form, the counselor asked patients the HIV risk questions using whatever wording and approach he felt would elicit the most reliable and valid information. The mean length of the interviews was essentially the same in the written-questionnaire (mean = 5.58 minutes, SD = 2.68 minutes) and audio-questionnaire (mean = 5.59 minutes, SD = 2.83 minutes) study groups ($P = .98$). The interviewer estimated that the written questionnaire also took an average of 6 minutes to complete. The audio questionnaire took 14 minutes to complete if the tape was not stopped or rewound.

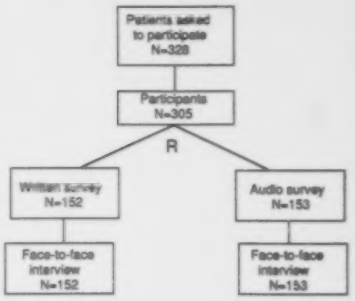


FIGURE 1—Study design.

For the purposes of this study, all 16 study variables were dichotomized and coded as "risk identified" vs "risk unidentified." Hence, categorical and continuous variables were simplified to two categories (i.e., sometimes, often, or always used nonprescription drugs to get high before or during sex vs never used nonprescription drugs to get high before or during sex). "Risk unidentified" was either a response indicating no risk or a missing response on that item. Missing responses were not excluded from the analyses because the mean percentage of missing responses across 16 items varied by risk assessment method. Hence, excluding missing responses introduced sample bias when comparisons were made between study groups. Also, including missing responses in the analysis appropriately attributed practical importance to missing data and eliminated the possibility of biasing groups with missing data toward risk identification. The approach allowed missing cases to be used in the analyses and allowed comparison of the number of identified risk factors by risk assessment method.

The chi-square test was used to compare the proportions of subjects reporting risk factors in the written and audio study groups. Student's t test was used to compare the mean number of risk factors reported in the written and audio study groups. McNemar's test for paired comparisons was used to compare propor-

TABLE 1—Characteristics of Study Patients,[a] by Study Group

	Study Group	
	Written (n = 152), %	Audio (n = 153), %
Female	41	39
Age, y		
≤20	23	25
21–29	53	50
≥30	24	25
Race		
White	19	18
Black	73	74
Hispanic	5	5
Other	3	3
Education		
<12 y	15	18
High school diploma	31	28
Post–high school education	44	38
College degree	11	16
Age at first intercourse, y		
≤13	25	26
14–15	31	29
≥16	44	45

[a]As reported in face-to-face interview.

TABLE 2—Number of Missing Cases in the Self-Administered Questionnaires (SAQs) That Face-to-Face Interviews (FTFIs) Identified as at Risk

	Written Study Group (n = 152)		Audio Study Group (n = 153)	
Risk Factor	No. Missing Cases in SAQs	No. Missing Identified as at Risk in FTFI	No. Missing Cases in SAQs	No. Missing Identified as at Risk in FTFI
History of sexually transmitted disease	40	12	47	3
Partner with HIV infection/AIDS	1	...	0	...
Multiple male partners	11	...	4	...
Multiple female partners	17	1	3	...
Sex with homosexual or bisexual man	12	...	3	...
One-night stand	1	...	2	1
Prostitution with male	8	...	2	...
Prostitution with female	5	...	1	...
Sex with intravenous drug user	3	...	0	...
Alcohol use with sex	1	...	1	...
Illicit drug use with sex	1	...	1	1
Shared intravenous needles	4	...	2	...
Unprotected vaginal sex with a steady partner	7	5	2	1
Unprotected vaginal sex with a nonsteady partner	11	1	0	...
Unprotected receptive anal sex with a steady partner	16	...	3	...
Unprotected receptive anal sex with a nonsteady partner	18	1	2	...

tions of subjects reporting risk factors on the questionnaires and interviews. Paired *t* tests were used to compare the mean number of risk factors reported on the questionnaires and interviews. Repeated-measures analysis of variance (ANOVA) was used to compare the mean number of risk factors identified in the written and audio questionnaires while controlling for interview responses. Multiple regression was used to determine whether patient characteristics were independently associated with reporting risk factors in either the questionnaire or the interview, but not in both these methods. Statistical significance was considered to be $P \leq .05$ except where otherwise indicated.

Results

The two study groups had similar demographic profiles (Table 1). The study population was 40% female and 73% Black, with a mean age of 25 years (SD = 6.37 years). Over half (56%) of study participants had more than a high school education.

The number of missing responses for each risk factor is shown in Table 2. The mean percentage of missing responses across 16 items varied by risk assessment method: 6.4 for the written questionnaires, 3.0 for the audio questionnaires, and 0 for the interviews. The number of missing responses was greater for the written questionnaires (mean = 1.01, SD = 1.44) than for the audio questionnaires (mean = 0.48, SD = 0.72) ($P \leq .0001$). Most patients who skipped a risk factor item in the questionnaire did *not* identify risk on that risk factor in the interview (87% of the missing responses in the written questionnaire and 92% in the audio questionnaire corresponded to no identified risk in the interview).

There were no differences in the number of risk factors reported in the written-questionnaire study group (mean = 3.43, SD = 2.03) compared with the audio-questionnaire study group (mean = 3.65, SD = 2.05) ($P = .35$; Table 3). However, responses to 2 of the 16 individual risk factor items differed between the written and audio questionnaires ($P < .05$): more patients in the audio study group than in the written study group reported having had a sexual partner who was known or suspected to be HIV-infected and having had unprotected vaginal intercourse with a nonsteady partner.

The interview responses of the written- and audio-questionnaire study groups to 16 individual risk factor items were compared to determine whether the study groups had similar risk profiles (Table 3). Two differences were detected: in the interviews, the written study group reported a history of sexually transmitted disease and having had a one-night stand more often than did the audio study group. Overall, the number of reported risk factors in the interviews was not different between the written (mean = 2.97, SD = 1.77) and audio (mean = 2.68, SD = 1.86) study groups ($P = .16$).

The percentages of discordant responses between the written questionnaires and the interviews and between the audio questionnaires and the interviews were determined for each risk factor (Table 4). The number of risk factors identified in the audio questionnaire and not in the interview (mean = 1.27,

TABLE 3—Percentage of Sexually Transmitted Disease Clinic Patients Reporting Risky Behavior in Self-Administered Questionnaires and Face-to-Face Interviews, by Study Group

Risk Factor	% Reporting Risk Behavior in Questionnaires			% Reporting Risk Behavior in Interviews		
	Written Study Group (n = 152)	Audio Study Group (n = 153)	Chi-Square *P*	Written Study Group (n = 152)	Audio Study Group (n = 153)	Chi-Square *P*
History of sexually transmitted disease	36.2	30.7	NS	37.5	22.9	.005
Partner with HIV infection/AIDS	1.0	4.6	.02[a]	1.3	0.7	NS[a]
Multiple male partners	18.4	18.3	NS	15.8	15.7	NS
Multiple female partners	38.8	34.0	NS	38.2	31.4	NS
Sex with homosexual or bisexual man	1.3	1.3	NS[a]	0.7	0.0	NS[a]
One-night stand	32.9	26.8	NS	32.2	22.2	.05
Prostitution with male	2.6	4.6	NS	2.0	3.9	NS[a]
Prostitution with female	7.2	3.3	NS	5.9	2.0	.09
Sex with intravenous drug user	4.0	3.3	NS	0.7	0.7	NS[a]
Alcohol use with sex	53.3	54.2	NS	46.1	44.4	NS
Illicit drug use with sex	11.2	11.8	NS	13.2	14.4	NS
Shared intravenous needles	2.6	2.6	NS[a]	1.3	0.7	NS[a]
Unprotected vaginal sex with a steady partner	66.5	75.2	.09	66.5	68.7	NS
Unprotected vaginal sex with a nonsteady partner	45.4	62.1	.003	32.2	37.3	NS
Unprotected receptive anal sex with a steady partner	13.8	19.6	NS	3.3	2.0	NS[a]
Unprotected receptive anal sex with a nonsteady partner	8.6	13.1	NS	0.7	0.7	NS[a]

Note. NS = not significant (chi-square *P* value of > .10).
[a]Not significant by Fisher's Exact Test.

SD = 1.16) was significantly greater than the number identified in the written questionnaire and not in the interview (mean = 0.99, SD = 1.16) ($P \leq .04$).

Paired comparisons of patient-reported risk factors between the questionnaire and the subsequent interview revealed differences. The number of risk factors reported by patients first in the written questionnaire (mean = 3.43, SD = 2.03) and then in the interview (mean = 2.97, SD = 1.77) differed significantly ($P \leq .0001$) (Table 4). Responses to 4 of 16 individual risk factor items compared between the written questionnaire and the interview were different: more patients reported in the written questionnaire that they used alcohol with sex ($P \leq .05$), had unprotected vaginal intercourse with a nonsteady partner ($P \leq .01$), had unprotectd receptive anal intercourse with a steady partner ($P \leq .001$), and had unprotected receptive anal intercourse with a nonsteady partner ($P \leq .05$).

The number of risk factors reported by patients first in the audio questionnaire (mean = 3.65, SD = 2.05) and then in the interview (mean = 2.68, SD = 1.86) also differed significantly ($P \leq .0001$). Responses to 6 of 16 individual items compared between the audio questionnaire and the interview were different: more patients reported in the audio questionnaire that they had a history of sexually transmitted disease ($P \leq .05$), had sex with someone who was known or suspected to be HIV-infected ($P \leq .05$), used alcohol with sex ($P \leq .001$), had unprotected vaginal intercourse with a nonsteady partner ($P \leq .001$), had unprotected anal intercourse with a steady partner ($P \leq .001$), and had unprotected anal intercourse with a nonsteady partner ($P \leq .001$). A repeated-measures ANOVA indicated that the difference between the mean numbers of risk factors reported in the audio questionnaires and in the interviews was greater than that between the mean numbers reported in the written questionnaires and in the interviews (mean audio = 3.65 ± 2.05; mean post-audio interview = 2.68 ± 1.86; mean written = 3.43 ± 2.03; mean post-written interview = 2.97 ± 1.77) ($P = .001$).

The number of discordant responses in which risk was identified in the questionnaire and not in the interview was not associated with patient characteristics including study group, age, race, gender, level of education, or age at first intercourse in multiple regression analysis. The number of discordant responses in which risk was identified in the interview and not in the questionnaire was independently, but weakly, associated with being in the written-questionnaire study group (standardized $\beta = -0.15$, $P \leq .008$), having less education (standardized $\beta = -0.18$, $P \leq .003$), and being male (standardized $\beta = -0.14$, $P \leq .01$).

Conclusions

This study suggests that an audio self-administered questionnaire has advantages over a written self-administered questionnaire for assessing patient HIV risk. This study also raises doubts that a face-to-face interview by an experienced interviewer is the optimal method for HIV risk assessment of all patients.

There were more missing responses in the written than in the audio questionnaires. Matching of responses between the written questionnaire and the interview revealed that many patients who skipped a risk factor item in the written questionnaire reported in the interview that they did not have that risk factor—except for responses about a history of sexually transmitted disease and unpro-

TABLE 4—Percentage of Sexually Transmitted Disease Clinic Patients Reporting Risky Behaviors Differently in Self-Administered Questionnaires (SAQs) and Face-to-Face Interviews (FTFIs), by Study Group

Risk Factor	Written Study Group (n = 152)		Audio Study Group (n = 153)	
	% Only SAQ Reported	% Only FTFI Reported	% Only SAQ Reported	% Only FTFI Reported
History of sexually transmitted disease	9.2	10.5	12.4	4.6*
Partner with HIV infection/AIDS	0.7	1.3	3.9	0.0*
Multiple male partners	3.9	1.3	3.3	0.7
Multiple female partners	4.6	3.9	5.2	2.6
Sex with homosexual or bisexual man	0.7	0.0	1.3	0.0
One-night stand	4.6	3.9	6.5	2.0
Prostitution with male	0.7	0.0	0.7	0.0
Prostitution with female	2.6	1.3	2.0	0.7
Sex with intravenous drug user	3.3	0.0	2.6	0.7
Alcohol use with sex	9.9	2.6*	10.5	0.7***
Illicit drug use with sex	2.6	4.6	2.6	5.2
Shared intravenous needles	2.6	1.3	2.0	0.0
Unprotected vaginal sex with a steady partner	11.8	11.8	15.0	8.5
Unprotected vaginal sex with a nonsteady partner	21.7	8.6**	28.8	3.9***
Unprotected receptive anal sex with a steady partner	11.8	1.3***	17.6	0.0***
Unprotected receptive anal sex with a nonsteady partner	8.6	0.7*	12.4	0.0***

*$P \leq .05$; **$P \leq .01$; ***$P \leq .001$; all by McNemar's test.

tected vaginal intercourse with a steady partner. It is unclear why a history of sexually transmitted disease was frequently unreported in the written questionnaire but was reported in the interview. For other questions, comparison of matched interview responses with missing responses in the written questionnaire suggested that patients sometimes skipped questions in the written questionnaire when they believed the questions did not apply to them because they usually did not report risk in the interview. The readability of complex questions may have been a problem for some patients. Compared with the written questionnaire, the audio questionnaire resulted in less skipping, perhaps because readability was less of a problem and attention to instructions requesting responses to all questions was greater when the audio method was used.

Although both types of questionnaires elicited the same number of reported risk factors when administered to randomized groups of patients, analysis of individual risk factor items suggested that the audio questionnaire identified more risk than the written questionnaire. On two individual risk factors—having had a sexual partner who was known or suspected to be HIV-infected and unprotected vaginal intercourse with nonsteady partners—significantly more patients reported risk in the audio questionnaire than in the written questionnaire. Interviews of these two groups demonstrated no differences on these risk factors. Furthermore, when discordant responses between the questionnaires and the interviews were analyzed, the audio questionnaires identified more risk factors than did the written questionnaires.

The educational level of this study population of patients at a sexually transmitted disease clinic was fairly high, indicating a high level of literacy. Also, the questionnaires were completed in private rather than in waiting rooms or nonprivate rooms in this study. These factors may have diminished the differences in reported risk between the written and audio questionnaires. Future research should determine whether the audio questionnaire is more effective in eliciting and identifying risk than the written questionnaire when administered to a population with a lower level of education and in waiting rooms rather than in private rooms.

More risk factors were identified in both the written and audio questionnaires than in the interviews. When rates of individual risk factors were compared among methods, more alcohol use with sex, more unprotected vaginal intercourse with nonsteady partners, and more anal intercourse with both steady and nonsteady partners were identified in both the written and audio questionnaires compared with the interviews. History of sexually transmitted disease and having had a partner with AIDS were also more often identified in the audio questionnaire than in the interview. These results indicated that both questionnaires identified more risk than the interview but that the audio questionnaire identified more risk compared with the interview than did the written questionnaire.

There were several possible explanations for the differences in responses observed between the questionnaires (written and audio) and the interviews. The order of administration of risk assessment methods (questionnaire followed by interview) may have influenced risk factor reporting. Self-administered questionnaires could have "primed" patients to conceal risky behaviors in the interviews, although it is just as plausible that they could have primed patients to reveal risky behaviors in the interviews. Face-to-face interviews might have been rushed due to patient impatience or interviewer concern about time and therefore might have been less thorough than the questionnaires. The interviews did take less time than the questionnaires. Nevertheless, concern over time would have been greater for postaudio interviews than for postwritten interviews due to the longer length of the audio questionnaire. Yet the interviews were the same length in both the written and audio study groups. Finally, the differences observed in risk reporting between the questionnaires and the interviews may have been at least in part attributable to the greater anonymity provided by the questionnaires.

A likely explanation for the differences in responses between the questionnaires and the interviews was that the experienced sexually transmitted disease clinic interviewer did not obtain accurate information for all risk factors. That interviewers, including physicians, have difficulty discussing patients' specific sexual practices has been observed in previous studies.[14,15] Although the interviewer was allowed to rephrase questions to encourage more revelation of risk and to reduce repetitiveness with the prior question-

naire, this rephrasing may have resulted in less direct questioning about sensitive behaviors, thus limiting patient responses. Also, patients may have been less revealing in the interview than in the questionnaires because they were uncomfortable with the interviewer.

We attempted to characterize patients who identified risk in the questionnaires but not in the interviews, and vice versa, to provide insight as to which types of individuals revealed more risk in which type of assessment. Significant findings were that the number of discordant responses in which risk was identified in the interviews but not in the questionnaires was weakly correlated with being in the written-questionnaire group, having less education, and being male. These findings suggest that the interviews identified more risk information after the written questionnaires than after the audio questionnaires, further confirming that audio questionnaires are more thorough than written questionnaires. Individuals with less education were probably less able to understand and complete written questionnaires compared with audio questionnaires. Finally, male more than female patients were more likely to reveal risk, and perhaps even to overreport risk, in the interviews compared with the questionnaires. That the interviewer was male may have contributed to this finding because match of the interviewer and patient on gender may affect interview responses.

A limitation of this study is that there is no way of determining which method of risk assessment provided the most valid responses. Because it seems more likely that patients would underreport rather than overreport their risky sexual practices to avoid embarrassment or disapproval, we assume that the method in which patients reported the greatest number of risk factors is the more valid. It was possible that patients erroneously reported risk factors because they misunderstood particular questions, especially in the written questionnaire, which depended on high literacy. For clinical purposes, the method that errs toward overreporting of risk might be preferred to one that errs toward underreporting of risk because diagnostic and educational follow-up is usually provided only for those patients for whom risk is identified. Overreporting would also be preferred in screening for blood donation.

Findings of this study support previous research suggesting that self-administered questionnaires may reduce barriers to HIV risk factor reporting found in face-to-face interviews,[12] although reading and comprehension difficulties can decrease the usefulness of questionnaires. Compared with a written questionnaire, an audio questionnaire has the potential advantages of being confidential in a waiting room, requiring minimal reading ability, and encouraging attention to every instruction, question, and question response. This study suggests that an audio questionnaire obtains more complete data and identifies more risk factors than a written questionnaire. Based on the findings of this study, further research is warranted as to whether assessing HIV risk with an audio questionnaire can overcome reading, attitudinal, and situational barriers to survey completion and accuracy. Also, this study emphasizes the need to further elucidate the relative strengths and weaknesses of written and audio questionnaires and interviews for HIV risk assessment. □

Acknowledgments

This study was supported by a grant from the Centers for Disease Control through a cooperative agreement with the Association of Teachers of Preventive Medicine.

We wish to thank Lynn Frank, James Matthews, Tina Clark, Eileen Burroughs, and Arthur Brown at the Montgomery County Health Department for their assistance in developing and implementing this project.

References

1. Saltzman SP, Stoddard AM, McCusker J, Moon MW, Mayer K. Reliability of self-reported sexual behavior risk factors for HIV infection in homosexual men. *Public Health Rep.* 1987;102:692–697.
2. Quinn TC, Glasser D, Cannon RO, et al. Human immunodeficiency virus infection among patients attending clinics for sexually transmitted disease. *N Engl J Med.* 1988;318:197–203.
3. Cannon RO, Schmid GP, Moore PS, Pappaioanou M. Human immunodeficiency virus (HIV) seroprevalence in persons attending STD clinics in the United States, 1985–1987. *Sex Transm Dis.* 1989;16:184–189.
4. Wasser SC, Aral SO, Reed DS, Bowen GS. Assessing behavioral risk for HIV infection in family-planning and STD clinics. *Sex Transm Dis.* 1989;16:178–183.
5. Locke SE, Kowaloff HB, Hoff RG, et al. Computer-based interview for screening blood donors for risk of HIV transmission. *JAMA.* 1992;268:1301–1305.
6. Padian NS, Shiboski SC, Jewell NP. Female-to-male transmission of human immunodeficiency virus. *JAMA.* 1991;266:1664–1667.
7. Lifson AR, O'Malley PM, Hessol NA, Buchbinder SP, Cannon L, Rutherford GW. HIV seroconversion in two homosexual men after receptive oral intercourse with ejaculation: implications for counseling concerning safe sexual practices. *Am J Public Health.* 1990;80:1509–1510.
8. Upchurch DM, Ray P, Reichart C, Celentano DD, Quinn T, Hook EW. Prevalence and patterns of condom use among patients attending a sexually transmitted disease clinic. *Sex Transm Dis.* 1992;19:175–180.
9. Coates RA, Soskolne CL, Calzavara L, et al. The reliability of sexual histories in AIDS-related research: evaluation of an interview-administered questionnaire. *Can J Public Health.* 1986;77:343–348.
10. Ekstrand ML, Coates TJ. Maintenance of safer sexual behaviors and predictors of risky sex: the San Francisco Men's Health Study. *Am J Public Health.* 1990;80:973–977.
11. McKusick L, Coates TJ, Morin SF, Pollack L, Hoff C. Longitudinal predictors of reductions in unprotected anal intercourse among gay men in San Francisco: the AIDS Behavioral Research Project. *Am J Public Health.* 1990;80:978–983.
12. Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. *Psychol Bull.* 1990;108:339–362.
13. Doak C. *Teaching Patients with Low Literacy Skills.* Philadelphia, Pa: JB Lippincott Co; 1985.
14. Johnson WT, Delamater JD. Response effects in sex surveys. *Public Opinion Q.* 1976;40:165–181.
15. Boekeloo BO, Marx ES, Kral AH, Coughlin SC, Bowman M, Rabin DL. Frequency and thoroughness of STD/HIV risk assessment by physicians in a high-risk metropolitan area. *Am J Public Health.* 1991;81:1645–1648.

(*Appendix on next page*)

APPENDIX—HIV Risk Assessment Questions

1. In the *past 12 months,* how many times have you been diagnosed with an STD or VD?
2. Have you ever had sex with someone who you know, or suspected, had a positive blood test for the AIDS virus?
3. *In the past 3 months,* with about how many different men did you have sex?
4. *In the past 3 months,* with about how many different women did you have sex?
5. At any time *in the past 3 months,* did you have sex with a man who had sex with other men?
6. *In the past 3 months,* have you had a "one night stand" with anyone?
7. At any time *in the past 3 months* did you give or receive money or drugs for sex with a boy or man?
8. At any time *in the past 3 months* did you give or receive money or drugs for sex with a girl or woman?
9. *In the past 3 months,* about how often did you use alcohol including liquor, beer, or wine of any type before or during sex?
10. *In the past 3 months,* about how often did you use nonprescription drugs to get high before or during sex?
11. *In the past 3 months,* have you had sex with anyone who injected drugs with needles?
12. Have you ever used a needle or syringe *after* someone else used it or may have used it?

The next set of questions ask about your condom use with a *steady partner*—your boyfriend, girlfriend, or spouse only. If you had sex with other people, these questions do not pertain to them. In the *past 3 months:*

(For men and women)

13. About how often was a condom used when your partner put his penis in your rectum?

(For women)

14. About how often was a condom used when your partner put his penis in your vagina?

(For men)

15. About how often was a condom used when you put your penis in your partner's vagina?
16. About how often was a condom used when you put your penis in your partner's (male or female) rectum?

The next set of questions ask about your condom use with a *nonsteady partner(s)*—people other than your boyfriend, girlfriend, or spouse. In the *past 3 months:*

(For men and women)

17. About how often was a condom used when a man put his penis in your rectum?

(For women)

18. About how often was a condom used when a man put his penis in your vagina?

(For men)

19. About how often was a condom used when you put your penis in your partner's vagina?
20. About how often was a condom used when you put your penis in your partner's (male or female) rectum?

HIV Vaccine Trials: Will Intravenous Drug Users Enroll?

Kathleen Meyers, MS, David S. Metzger, PhD, Helen Navaline, BA, George E. Woody, MD, and A. Thomas McLellan, PhD

ABSTRACT

Objectives. The purpose of this study was to assess the willingness of intravenous drug users to participate in a preventive human immunodeficiency virus (HIV) vaccine efficacy trial.

Methods. Of the 347 intravenous drug users in methadone treatment who were approached for participation, 257 completed a battery of self-administered questionnaires assessing risk behaviors, interest in vaccine trials, and other vaccine-related information. Data from 16 known seropositives and 1 inconsistent responder were dropped from analyses (n = 240).

Results. Fifty-two percent of the subjects expressed a willingness to be one of the first individuals to participate in a preventive HIV vaccine efficacy trial. Subjects who had recently shared needles or works and subjects who trusted the government to ensure vaccine safety were both twice as likely to report interest in participation. Twenty-two percent of subjects reported that they would increase needle sharing if vaccinated. Thirty percent did not know what a vaccine was.

Conclusions. These findings suggest that some in-treatment intravenous drug users would volunteer for a preventive HIV vaccine efficacy trial. Education and counseling will be required to ensure that subjects fully understand the trial's purposes, methods, risks and benefits. (*Am J Public Health.* 1994;84:761–766)

Introduction

Despite impressive risk reduction resulting from behavioral and educational interventions launched to halt the spread of human immunodeficiency virus (HIV) infection, sizable portions of both the homosexual/bisexual and intravenous drug using populations continue to engage in, or have relapsed into, behaviors that place them at risk of acquiring the virus.[1–4] Of parallel concern is the apparent absence of meaningful risk change among the large number of individuals who do not view themselves at particular risk (e.g., adolescents, adult heterosexuals).[5–7] In fact, the acquired immunodeficiency syndrome (AIDS) has emerged as a leading cause of death among US adults less than 45 years of age and US children 1 to 5 years old.[8] Thus, there is an urgent need for the development and testing of preventive HIV vaccines.

Currently, a number of experimental HIV vaccines have entered phase I and II clinical trials.[9–11] As this work progresses and preparation for phase III trials begins, epidemiological, ethical, and public health considerations are being cooperatively addressed by national and international groups.[12–16] At the same time, a critical practical question is the willingness of at-risk individuals to participate in the upcoming vaccine trials. To date, little work has been undertaken to delineate the issues associated with recruitment and retention of subjects for HIV vaccine trials.

An initial step in testing the efficacy of candidate HIV vaccines is gaining access to populations at risk of HIV infection. Under a cooperative initiative among the National Institute of Allergy and Infectious Diseases (NIAID), the National Institute on Drug Abuse (NIDA), and the Centers for Disease Control, homosexual/bisexual men, intravenous drug users, and other at-risk individuals are being targeted for participation in these efficacy trials.[12] The active participation of gay men in helping to formulate the HIV research and treatment agenda suggests that this community recognizes the importance of vaccine trials, although their willingness to be participants awaits empirical examination. There is less certainty regarding the willingness of intravenous drug users to participate in vaccine trials, since they are not well organized and tend to be a disenfranchised group.[17] Consequently, it is important to develop some indication of intravenous drug users' willingness to volunteer for such a study. Most important is determination of whether seronegative subjects who are at the greatest risk of infection will be willing to participate. These assessments will have direct bearing on the estimate of sample sizes needed for vaccine trials. For example, if those at greatest risk for seroconversion are unwilling to enroll, sample sizes for vaccine trials will need to be much larger than if such individuals are willing to participate.

With support from the NIAID/NIDA vaccine preparedness initiative, we had the unique opportunity to rapidly collect preliminary data on the willingness of intravenous drug users to participate in efficacy trials for HIV preventive vaccines.

The authors are with the Center for Studies of Addiction, University of Pennsylvania/Philadelphia Department of Veterans Affairs.

Requests for reprints should be sent to Kathleen Meyers, MS, University of Pennsylvania/VA Center for Studies of Addiction, PVAMC Bldg #7, University and Woodland Aves, Philadelphia, PA 19104.

This paper was accepted August 3, 1993.

As part of recruitment procedures for a new cohort of subjects participating in a longitudinal study of risk behaviors and HIV infection among intravenous drug users in Philadelphia,[18] methadone patients were questioned about their willingness to participate in early vaccine trials. Subjects were also briefly questioned regarding their knowledge of vaccines, trust in government, and other issues considered important in gaining a preliminary understanding of the factors associated with trial participation. Methadone patients not only represent the target population of intravenous drug users in treatment but provide access to their out-of-treatment intravenous drug using associates, a group at particularly high risk. Therefore, they constitute a readily available sample from which we can gain some early insight regarding intravenous drug users' willingness to enroll in HIV vaccine trials and an important link to the out-of-treatment community.

This paper presents self-report data from a brief opinion questionnaire about these issues among 240 intravenous drug users. It explores the feasibility of recruiting such individuals for clinical vaccine trials, examines whether seronegative subjects at greatest risk of HIV infection report willingness to participate in HIV vaccine trials, and discusses issues related to their ability to provide informed consent for such trials.

Methods

Procedures

The study sample was composed of methadone maintenance patients from the Girard Methadone Program located in north-central Philadelphia. All eligible methadone patients were approached for screening as part of recruitment procedures for a longitudinal study of HIV infection and associated risk behaviors.[18] Of those screened (n = 347), 257 (74%) agreed to complete our initial assessment.

After providing informed consent, subjects completed a battery of self-administered risk assessment questionnaires in closely supervised groups of 5 to 10. Included in the battery was a one-page vaccine opinion questionnaire developed specifically for this project. Throughout questionnaire completion, research staff members were present to answer questions, assist those with reading difficulty, and screen completed questionnaires for missing data or inconsistent responses. Subjects were paid $5 for their participation.

Subjects

Of the 257 subjects who completed questionnaires, 16 reported being HIV positive and 1 provided inconsistent responses. The data from these 17 subjects were deleted from further analyses, resulting in a final sample of 240 subjects. Subjects ranged in age from 20 to 64 years, with a mean of 39.1 years (SD = 7.0). The sample was ethnically diverse (44% African American, 39% Caucasian, and 17% Latino). Seventy percent of the sample was male. The median monthly income was $246, with most subjects (80%) receiving public assistance as their primary source of income. Fifty-nine percent had a high school diploma or general equivalency diploma. The median length of time in the current methadone program was 1.5 years.

Measures

The Vaccine Opinion Questionnaire was developed for this project on the basis of input from both intravenous drug users and research staff. While the assessment battery was being developed, we asked subjects in our original cohort to comment on vaccine research and what issues might be important for participation from the perspective of potential participants. We also asked our research staff to suggest issues that should be examined. These staff members have extensive contact with intravenous drug users, and they were encouraged to report issues and concerns that existed among the intravenous drug using community in regard to vaccines, treatment, and related issues. On the basis of this information, a brief self-report questionnaire was designed to identify some of the potentially important factors associated with trial participation. The resultant questionnaire was composed of 13 statements with which the subject could either agree or disagree and 3 multiple-choice questions. The questions, which were grouped post hoc, covered HIV-related issues, vaccine histories of both the respondent and his or her children, trust in government, vaccination knowledge, potential behavioral change after participation, and willingness to participate.

Formal reliability and validity testing was not conducted because our opportunity to examine this cohort was time limited. Therefore, we took care to maximize the face validity of the items constituting the questionnaire. For example, we used simple language that would be understood by this population. We also monitored all subjects during and after administration of the questionnaire, answering questions and clarifying inconsistencies. For these reasons, we are confident of the face validity of this instrument. Obviously, the only conclusive method of testing the predictive validity of the instrument would be to correlate answers on the questionnaire with actual participation in a vaccine trial. We will complete this formal study when the opportunity arises.

The Risk Assessment Battery,[19] an instrument composed of 38 closed-ended items that cover frequency and route of administration of various substances, needle sharing and cleaning, and high-risk sexual behavior, including sexual orientation, condom use, and exchange of sex for drugs or money, was used in collecting risk-associated behavioral data. Preliminary validity data show the instrument to have good predictive validity for HIV seroconversion.[19] The Risk Assessment Project Questionnaire yielded demographic and descriptive information about the subjects' current social conditions, drug treatment, history of criminal behavior, and legal history. The entire test battery took approximately 30 minutes for most subjects to complete. A considerable body of literature now exists supporting the validity and reliability of self-report measures within this population.[20,21]

Results

Trust

As can be seen in Table 1, the subjects expressed considerable distrust in the government. Fifty-three percent of the subjects believed that the health department was not doing all that it could to stop the spread of AIDS, and 32% believed that a cure for AIDS was available but being kept from the public. Thirty-six percent endorsed the belief that HIV is a man-made virus created to get rid of "certain groups" of people. Specifically related to trust in vaccine trials, 48% of the subjects did not trust the government to ensure vaccine safety prior to trials. This is particularly troublesome for recruitment, given that safety issues emerged as the most important issue reported by respondents (42%) regarding their participation.

Vaccine Knowledge/Experience

Responses indicated that the subjects were in need of education regarding vaccines. Thirty percent of the sample

were unsure what a vaccine was, and 41% did not know that a vaccine could prevent disease acquisition. This apparent lack of knowledge existed despite high rates of vaccination for other communicable diseases: 84% reported having been vaccinated for measles, mumps, or polio, and 93% of those subjects with children (n = 192) indicated that their children had been vaccinated. A higher proportion of those who admitted not really knowing what a vaccine was than of those who reported such knowledge did not have a high school education or a general equivalency diploma (55% vs 35%; $\chi^2 = 8.3$, $df = 1, P < .005$).

TABLE 1—Subjects' Opinions Regarding Vaccines and Related Issues, by Topic Area (n = 240)

	% Agreeing or Answering Yes
Trust	
The health department is doing all it can to stop the spread of AIDS.	47
There is a cure for AIDS, but the government is keeping it from the public.	32
I trust the government to make sure that vaccines they want to test are safe before they test them on people.	52
HIV is a manmade virus that was created to get rid of certain groups of people.	36
Vaccine knowledge/experience	
A vaccine can protect you from getting a disease.	59
A vaccine does not help someone infected with a disease.	48
Getting vaccinated can be risky.	48
I am not sure what a vaccine is.	30
Have you ever received a vaccine for a disease like measles, mumps, or polio?	84
Were your children vaccinated for things like measles, mumps, or polio?	93
Behavioral impact	
If I was sure I couldn't get AIDS, I'd probably share needles more often.	22
Trial participation willingness	
I would be willing to be one of the first people to try an HIV vaccine.	52
I would be willing to try a vaccine for HIV even if there was a slim chance of getting sick from it.	39
What would be most important for you to try a vaccine for HIV if one were available?	
The vaccine was safe.	42
I would be guaranteed medical care for any problems related to the vaccine.	20
I would have the opportunity to be protected from getting HIV.	35
I would receive money for trying the vaccine.	3

Behavioral Impact

Regarding the potential impact of vaccine trial participation on high-risk behaviors, 22% of the subjects thought that they might share needles more often if they were guaranteed protection from HIV infection. This reported potential for increased risk behaviors was more common among current needle sharers (40%) than among those not sharing (17%) ($\chi^2 = 14.7$, $df = 1$, $P < .001$), but it was not more common among those who reported willingness to participate in the trials. Whether these subjects would attempt to determine their vaccination status was not examined.

Participation Willingness

Despite substantial mistrust and lack of knowledge, a slight majority (52%) of the subjects expressed a willingness to be "one of the first" individuals to participate in an HIV vaccine trial, and 39% reported that they would be trial participants even if there was a slim chance of getting sick as a result of vaccination. When subjects were questioned about the factors that would be most likely to influence their decision to enroll, safety issues emerged as most important, followed by the opportunity to be protected from HIV infection, guaranteed medical care for any vaccine-related problems, and monetary reimbursement for trial participation.

Associated Willingness Factors

To explore factors associated with willingness to participate, subjects were categorized according to their answers on that question. As can be seen in Table 2, those willing to be part of a vaccine trial were demographically similar to those who were unwilling to be part of a trial. However, willing subjects had spent significantly less time in the current methadone program ($t = 1.98$, $df = 236$, $P < .05$) and were more likely to report current health problems ($\chi^2 = 5.1$, $df = 1$, $P < .05$) than unwilling subjects. Not surprisingly, those willing to be vaccine trial participants were proportionately more likely to believe that the government would ensure vaccine safety prior to testing ($\chi^2 = 5.8$, $df = 1$, $P < .05$), although the groups were proportionately similar on other trust issues. Surprisingly, basic knowledge of vaccines, particularly being unsure of what a vaccine was and knowing that vaccines can be protective, was unrelated to participation willingness.

As stated, an important focus of this study was an exploration of whether subjects engaging in risk behaviors were willing to enroll in vaccine trials. To this end, subjects were grouped according to their risk status on HIV-associated drug behaviors (e.g., shared needles/did not share needles in the prior 6 months) and HIV-associated sexual behaviors (e.g., no or irregular condom use with two or more sexual partners in the prior 6 months). The chi-square statistic was used in conducting a univariate analysis of the proportion of willing participants within these groups.

Significantly more of those who perceived themselves as being at risk of HIV exposure than of those who did not perceive themselves as at risk reported willingness to be one of the first to try a vaccine (59% vs 41%; $\chi^2 = 7.3$, $df = 1$, $P < .01$). Similarly, willingness to try a vaccine was significantly higher among those who recently shared needles than among those who did not share needles (66% vs 48%; $\chi^2 = 6.2$, $df = 1$, $P < .01$). Willingness was also more common among those who engaged in other risky drug using behaviors such as sharing a cooker, cotton, or rinse water in the previous 6 months (62% among those who shared vs

TABLE 2—Factors Associated with Willingness to Participate in HIV Vaccine Trial

	Willing Subjects (n = 125)	Unwilling Subjects (n = 113)
Demographics		
Male, %	66	74
Race, %		
African American	41	48
Caucasian	41	38
Latino	18	12
Mean age, y (SD)	39.5 (6.9)	38.9 (7.2)
High school diploma or GED[a]	60	58
Employed, %	20	14
Treatment-related issues		
Mean no. of months in methadone treatment (SD)	24.3 (33.5)	33.2 (35.3)
Health problems reported,* %	48	34
Trust in government, %		
Believe government tests only those vaccines that are safe*	60	44
Believe health department is doing all it can to stop the spread of AIDS	46	47
Believe there is a cure for AIDS but the government is keeping it from the public	33	33
Knowledge and experience, %		
Believe vaccines are protective	61	58
Not sure what a vaccine is	27	34
Personal vaccination history	89	80

[a]General equivalency diploma.
*$P \leq .05$.

TABLE 3—Logistic Regression Model of Predictors of Willingness to Participate in HIV Vaccine Trials

	Willing Subjects	Odds Ratio	95% Confidence Interval
Demographic variables			
Male (n = 165), %	50	1.5	0.80, 2.8
Minority (n = 144), %	51	1.0	0.56, 1.9
Mean age, y	39.5	1.0	0.98, 1.1
Risk behaviors, %			
Share needles/works (n = 110)	61	2.1	1.2, 3.7*
Have unprotected sex with multiple partners (n = 45)	58	1.2	0.58, 2.5
Trust, %			
Trust government to ensure vaccine safety (n = 121)	60	2.0	1.1, 3.7*
Believe government is doing everything possible to stop the spread of AIDS (n = 110)	52	0.9	0.49, 1.6
Believe HIV is not man-made (n = 151)	53	0.9	0.46, 1.7
Believe government is not keeping cure from the public (n = 158)	53	0.8	0.44, 1.6
Vaccine knowledge, %			
Sure what a vaccine is (n = 166)	55	1.6	0.84, 3.0
Believe vaccines are protective (n = 139)	53	1.0	0.54, 1.7
Believe vaccines are therapeutic (n = 120)	49	0.7	0.39, 1.2

*$P < .05$.

45% among those who did not share; $\chi^2 = 6.5, df = 1, P < .01$).

Willingness was unrelated to our measures of risky sexual behaviors: 58% vs 51% for those engaging in and those not engaging in unprotected intercourse with multiple (two or more in past 6 months) sexual partners ($\chi^2 = 0.6$, not significant* [NS]) and 63% vs 51% for those engaging in and those not engaging in sex for drugs or money in the prior 6 months ($\chi^2 = 1.9, df = 1$, NS).

Logistic regression was used to simultaneously adjust for age, sex, and race as well as the potential predictors of willingness to participate in vaccine efficacy trials: engagement in high-risk behaviors, trust in government, and vaccine knowledge. As can be seen in Table 3, subjects who shared needles or works were twice as likely to report willingness to participate in trials, as were those who reported that they trusted the government to ensure vaccine safety prior to trials. Demographic variables, sexual HIV-risk-associated behaviors, trust issues other than safety, and knowledge of vaccines were not significantly associated with reported willingness to participate in vaccine trials.

Discussion

These data from our initial questionnaire suggest that some in-treatment intravenous drug users would volunteer for HIV vaccine trials. Importantly, those in-treatment users who continued to place themselves at risk for HIV infection were significantly more likely to report that they would be one of the first to try a vaccine than those who engaged in safer behaviors. If these findings are confirmed and high-risk intravenous drug users are truly willing to enroll, the number of subjects necessary for efficacy vaccine trials could be minimized.

However, just under half of the intravenous drug users questioned expressed an unwillingness to enroll in vaccine trials. It was clear that a substantial lack of knowledge about vaccines and distrust in the government were evident among this group, and these factors will have to be addressed if participation rates are to be maximized. The data reported here suggest that mistrust of the government regarding vaccine safety is a key mediating variable in the recruitment of intravenous drug users for HIV vaccine efficacy trials. Efforts to combat this lack of trust and provide safety data should guide subject recruitment and trial imple-

mentation from the outset. In addition to doing everything possible to ensure vaccine safety, it may be helpful to select field workers who are part of the intravenous drug using community, rather than governmental officials, to recruit and engage at-risk users.

While this paper has reported on the responses of in-treatment intravenous drug users, it will also be important for HIV vaccine trials to recruit and retain out-of-treatment users. In our ongoing study of HIV infection, we have successfully used our in-treatment cohort to refer us to out-of-treatment subjects. Our findings indicate that out-of-treatment intravenous drug users engage in substantially higher levels of risk behaviors and have higher seroconversion rates than their in-treatment counterparts.[18] Both of these are important considerations for vaccine trial participation.

This study showed no significant difference in willingness to participate among the racial groups. However, since a high proportion of intravenous drug users are African Americans and Latinos, sensitivity and responsiveness to minority issues will be critical in vaccine trials. As noted by El-Sadr and Capps,[22] minority recruits may be an especially distrustful group, considering the past patterns of insensitivity during clinical trials exemplified by the legacy of the Tuskegee Syphilis Study.

One of the most important issues that will need to be addressed prior to vaccine trials is that of *informed* consent. Although knowledge of vaccines as measured here was unrelated to participation willingness, the reported lack of general knowledge regarding vaccines is alarming. Approximately one third of our subjects admitted to not really knowing what a vaccine was, and more than half did not know that there are preventive and therapeutic vaccines. Of course, it may be that this is representative of the knowledge base of nonusers from the same socioeconomic strata. Despite this apparent lack of basic knowledge, many subjects agreed to consent to trial participation without adequate comprehension of the issues. This scenario will obviously test our ability to ensure that subjects know what vaccines are and fully understand the trial (e.g., placebo-controlled issues) and its uncertainties (e.g., potential risks, efficacy uncertainties) so that they can make truly informed and knowledgeable decisions regarding enrollment. This becomes an even more important issue given the educational level of this group of subjects. Clear and concise information regarding the type of vaccine (e.g., subunit vs live vectors) to be tested in each trial, the research method (e.g., placebo-controlled study), and the consequences of participation (e.g., vaccine-induced seroconversion) will need to be provided in a way that ensures comprehension. The ability to engage subjects in an open, honest dialogue in order to dispel myths, allay fears, and candidly address the type and amount of risk the trial entails will undoubtedly influence enrollment. Included should be a frank discussion of compensation issues, in the event that subjects sustain harm, as well as the potential for discrimination after an immune response resulting from vaccination. In addition, it may be necessary to include a brief factual quiz regarding vaccine participation with a "cutoff score" as part of subject consent procedures. This may be the only definitive method of ensuring truly informed consent in future trials.

As subjects learn more of the specifics of these proposed vaccine trials, willingness will undoubtedly change, although the direction of the change is unknown. In an effort to better understand potential participation issues, we are currently using both interviews and self-report measures to assess reported willingness to participate under conditions that will approximate actual trials (e.g., randomized, placebo-controlled trials). Information delivery methods (e.g., videotapes) and the individuals who disseminate such information (e.g., other intravenous drug users, health officials) will probably vary in efficacy; this area requires study.

An important ethical consideration for recruitment is remuneration for participation. High monetary rewards offered to economically disadvantaged groups can be considered enticing to the point of coercion. Thus, it is particularly important to identify the nonmonetary issues that are most strongly associated with willingness to participate. This will help to identify factors that can attract potential participants based only on scientific and humanitarian issues. Although financial incentives are important for study compliance, they must not become the sole reason for enrollment.

Although few subjects reported the possibility of an increase in risk behaviors, the impact of vaccine trial participation on a subject's practice of risk behaviors requires further attention. Relapse into unsafe behaviors as a result of a "sense of safety" through vaccine trial participation would be problematic not only if the initial vaccines are less than 100% effective, but also in light of other consequences (e.g., hepatitis, sexually transmitted diseases). The ability of intravenous drug users to determine whether or not they received a placebo or vaccine, as well as their interest in doing so, is currently unknown and under investigation. Further exploration of other key vaccine trial issues (e.g., subjects' knowledge of placebo issues, vaccine-induced seroconversion) is an ongoing aim of our current work.

While the data reported here provide some preliminary insight into the willingness of intravenous drug users to participate in HIV vaccine efficacy trials, the only true measure of the validity of these responses will be whether those who say they will participate actually enroll in efficacy trials. We are currently in the process of assessing the degree of stability of a subject's stated participation willingness over the course of the preparedness work. If changes in willingness occur over time, we will ascertain whether any factors correspond to this fluctuation. For example, changes in willingness may be related to changes in risk behaviors. Subjects may be more likely to consider participation in vaccine efficacy trials during periods of high-risk behaviors and less likely to consider participation during safer periods. Whether this phenomenon exists at all and whether it is consistent across categories of risk behaviors will be the subject of future reports.

From these preliminary data, it is clear that the recruitment of intravenous drug users into HIV vaccine efficacy trials will require a sustained cooperative effort on the part of the vaccine and drug research and treatment communities. Together, these communities must determine appropriate incentives and develop and disseminate appropriate educational materials that can provide accurate descriptions of the study and understandable information on potential trial risks. The way in which these issues are addressed will undoubtedly form the cornerstone of recruitment efforts. As vaccine efficacy trials approach, we must quickly address the above issues and our understanding of the factors associated with enrollment if we are to recruit and engage the intravenous drug–using community. □

Acknowledgments

This work was supported by National Institute on Drug Abuse grant R01-DA05593, with supplemental support from the National Institute of Allergy and Infectious Diseases (interagency memorandum of agreement Y01-AI-20025-01).

Portions of this paper were presented at the 55th Annual Scientific Meeting of the College on Problems of Drug Dependence, Inc, Toronto, Ontario, Canada.

We wish to thank Amy R. Sheon and the reviewers for comments and Maureen Burke, Rich Davis, Suzette Dyanick, Ray Incmikoski, Shoshana Waskow, Joseph Williams, and the staff of the Goldman Clinic, North Philadelphia Health Systems, for their assistance in the completion of the study.

References

1. Metzger DS, Woody GE, DePhilippis D, et al. Risk factors for needle sharing among methadone treated patients. *Am J Psychiatry.* 1991;148:636–640.
2. Calsyn DA, Saxon AJ, Wells EA, Greenberg DM. Longitudinal sexual behavior changes in injecting drug users. *AIDS.* 1992;6:1207–1211.
3. McCusker J, Stoddard AM, McDonald M, Zapka JG, Mayer KH. Maintenance of behavioral change in a cohort of homosexually active men. *AIDS.* 1992;6:861–868.
4. Zule WA. Risk and reciprocity: HIV and the injecting drug user. *J Psychoactive Drugs.* 1992;24:243–249.
5. Centers for Disease Control. Sexual behavior among high school students. *MMWR.* 1992;40:885–888.
6. Centers for Disease Control. Selected behaviors that increase risk of HIV infection among high school students—US, 1990. *MMWR.* 1992;41:231–240.
7. Holmes KK, Karon JM, Kreiss J. The increasing frequency of heterosexually acquired AIDS in the United States, 1983–1988. *Am J Public Health.* 1990;80:858–863.
8. Centers for Disease Control. Mortality attributable to HIV infection/AIDS—US, 1981–1990. *MMWR.* 1991;40:41–44.
9. Koff WC, Hoth DF. Development and testing of AIDS vaccines. *Science.* 1988;241:426–432.
10. Koff WC, Fauci AS. Human trials of AIDS vaccines: current status and future directions. *AIDS.* 1989;3:S125–S129.
11. Karzon DT, Bolognesi DP, Koff WC. Development of a vaccine for the prevention of AIDS: a critical appraisal. *Vaccine.* 1992;10:1039–1052.
12. Vermund SH, Fischer RD, Hoff R, et al. Preparing for HIV vaccine efficacy trials: partnerships and challenges. *AIDS Res Human Retroviruses.* 1993;9S:S127–S132.
13. Enel P, Charrel J, Reviron D, Manuel C, SanMarco JL. Ethical problems raised by anti-vaccine vaccination. *Eur J Epidemiol.* 1991;7:147–153.
14. Esparza J, Osmanov S, Kallings LO, Wigzell H. Planning for HIV vaccine trials: the World Health Organization perspective. *AIDS.* 1991;5:S159–S163.
15. Smith PG, Hayes RJ, Mulder DW. Epidemiological and public health considerations in the design of HIV vaccine trials. *AIDS.* 1991;5:S105–S111.
16. Mariner WK. Why clinical trials of AIDS vaccines are premature. *Am J Public Health.* 1989;79:86–91.
17. Coates TJ, Stall RD, Kegeles SM, Lo B, Morin SF, McKusick L. AIDS antibody testing: will it stop the AIDS epidemic? Will it help people infected with HIV? *Am Psychol.* 1988;43:859–864.
18. Metzger DS, Woody GE, McLellan AT, et al. Human immunodeficiency virus seroconversion among in- and out-of-treatment intravenous drug users: an 18-month, prospective follow-up. *J AIDS.* 1993;6:1049–1056.
19. Metzger DS, Woody GE, DePhilippis D, et al. The impact of HIV testing on risk for AIDS behaviors. In: Harris L, ed. *Problems of Drug Dependence 1991: Proceedings of the 53rd Annual Scientific Meeting, The Committee on Problems of Drug Dependence, Inc.* Rockville, Md: National Institute on Drug Abuse; 1992:297–298. NIDA Research Monograph 119.
20. Magura S, Goldsmith D, Casriel C, Goldstein PJ, Lipton DS. The validity of methadone clients' self-reported drug use. *Int J Addict.* 1987;22:727–749.
21. Samuels JF, Vlahiv D, Anthony JC, Chaisson RE. Measurement of HIV risk behaviors among intravenous drug users. *Br J Addict.* 1992;87:417–428.
22. El-Sadr W, Capps L. The challenge of minority recruits in clinical trials for AIDS. *JAMA.* 1992;267:954–957.

A Community-Based Heart Disease Intervention: Predictors of Change

Marilyn A. Winkleby, PhD, June A. Flora, PhD, and Helena C. Kraemer, PhD

ABSTRACT

Objectives. This paper presents a prospective examination of sociodemographic, psychosocial, and physiologic characteristics associated with positive change in cardiovascular disease risk factors during a 6-year multiple risk factor intervention study.

Methods. Data are presented on 221 women and 190 men (aged 25 through 74 years) who participated in four cohort surveys (1979 through 1985). A signal detection model was used to identify baseline variables that best divide the sample into subgroups on the basis of the probability of positive change in a composite risk factor score.

Results. Sixty-nine percent of the respondents showed a positive change in risk factor score during the intervention. The subgroup with the highest proportion of positive changers (83%) was composed of older adults (>55 years) with the highest perceived risk, highest health media use, and highest blood pressure and cholesterol levels. The subgroup with the lowest proportion of positive changers (42%) was the least educated, was the most likely to be Hispanic, and had the lowest health knowledge and self-efficacy scores.

Conclusions. The differing composition of subgroups who respond or do not respond to community cardiovascular disease interventions illustrates the need to develop specific interventions that target different age, socioeconomic, and cultural subgroups. (*Am J Public Health.* 1994; 84:767–772)

Introduction

The Stanford Five-City Project is a long-term field trial designed to test whether a comprehensive program of community organization and health education produces favorable changes in cardiovascular disease risk.[1] Changes in overall risk factors have been reported[2] and indicate a beneficial educational effect, especially with regard to blood pressure[3] and smoking.[4]

In this paper, we present prospective data on respondents from the two treatment cities. We identify those who showed favorable risk factor changes and compare their baseline sociodemographic, psychosocial, and physiologic characteristics with those of respondents who showed unfavorable changes. This procedure allows for the identification of subgroups who are most likely to respond to a community intervention like the Five-City Project, as well as subgroups who deserve special attention because of their lack of positive change.

Because intervention programs increasingly involve the targeting of multiple risk factors[1,5,6] and because the primary goal of community cardiovascular disease intervention studies is overall improvement in the community's health, we used a summary measure that incorporates multiple cardiovascular disease risk factors as our indicator of change. Using a signal detection model designed to develop optimal sequential rules,[7] we identified baseline health-related variables that best divide the sample into subgroups on the basis of the probability of positive change in risk factor status.

Our selected variables reflect the theoretical bases of the Project's education program: social learning theory,[8,9] the communication-persuasion model,[10] the social marketing model,[11] and the theory of reasoned action.[12] We evaluate knowledge and attitudes about cardiovascular disease that often accompany behavior change[13]; perceived self-efficacy about diet and exercise, which has been shown to be important in predicting behavior change[14]; perceived cardiovascular disease risk, which has been shown to be important in understanding how individuals change[15]; and health media use, which has been shown to be related to exposure to the Five-City Project campaign.[16]

Methods

The Five-City Project drew subjects from two treatment and two control cities in Northern California, ranging in population size from 35 000 to 145 000 residents (a fifth city was monitored for morbidity and mortality, but not for risk factor change). To assess change in risk factors, independent cross-sectional surveys of randomly selected households and repeated surveys of a cohort were conducted. All persons aged 12 through 74 years were eligible to participate and were invited to attend survey centers located in each community. Detailed descriptions of the study design and methodology have been published previously.[1,17,18]

This analysis focuses on adults aged 25 through 74 years living in the two treatment cities who participated in at least the baseline cohort survey (1979/80) and the final cohort survey (1984/85).

The authors are with the Stanford Center for Research in Disease Prevention, Stanford University School of Medicine, Palo Alto, Calif.

Requests for reprints should be sent to Marilyn A. Winkleby, PhD, Stanford Center for Research in Disease Prevention, Stanford University School of Medicine, 1000 Welch Rd, Palo Alto, CA 94304-1885.

This paper was accepted August 3, 1993.

When these subjects also participated in either or both of the interim cohort surveys (1980/81 and 1982/83), interim data were also used. Eighty-seven percent of those who participated in the baseline and final surveys also participated in both of the interim surveys.

The Education Intervention

The 6-year education intervention targeted all residents in the treatment communities and involved a multiple risk factor strategy delivered through multiple educational channels.[19] Mass media (television and radio) programs formed a major proportion of the intervention. Print media (newspaper health columns and pamphlets) were delivered through direct mail, at work sites, and through medical care givers.[20] Special emphasis was given to designing materials for those with low literacy levels and disseminating programs to low-income Spanish-speaking populations, particularly via radio.

Outcome Variable

As an overall measure of change in cardiovascular disease risk factors, we chose change in a composite risk factor function (based on the Framingham Study) that provides an estimate of all-cause mortality risk per 1000 persons in 10 years.[21] The risk function, which is sex-specific, is a combination of cigarettes per day, total plasma cholesterol, mean diastolic blood pressure, and pulse (men only). Although we recognize that an individual's response may be specific to one component of the risk score (e.g., smoking) but not to another component (e.g., cholesterol), our objective was to select an outcome indicative of overall improvement in health. Furthermore, a single outcome variable avoids problems of collinearity and multiple-hypothesis testing. It thus provides a more powerful method of assessing the simultaneous effects of multiple risk factors than do analytic methods that become impractical when separate variables are used.

Information on variables used in the risk assessment was obtained by experienced nurses and laboratory technicians. Respondents were asked if they had smoked any cigarettes in the past week. If no cigarettes were smoked, respondents received a code of zero; if one or more cigarettes were smoked, they received a code reflecting the number of cigarettes smoked per day. Although no biochemical measures were used to validate the number of cigarettes smoked per day, previous analyses have shown a high degree of accuracy between self-reported and biochemical measures for overall smoking status.[4,22] Total plasma cholesterol was derived from nonfasting venous samples and analyzed by standard methods established by the Lipid Research Clinic Program.[23] Three blood pressure measurements were taken on the right arm with a semiautomatic recorder (Sphygmetrics SR-2 automatic blood pressure recorder) that minimizes observer bias.[17] The mean of the second and third diastolic readings were used for analyses. Resting pulse was determined by direct palpation for 1 minute.

We calculated respondents' composite risk factor functions at each survey time from the values of their risk factors at that time. (The mean baseline risk score was 3.5 [SD = 1.2].) Age was held constant for each individual at his or her baseline age, allowing for the examination of change in risk due to factors targeted by the intervention. After calculating a respondent's risk function values at each survey, we employed a random regression model to fit a unique slope for each respondent.[24] On the basis of the value of their individual slopes, individuals were divided into two groups, one representing positive changers (those with a positive slope, indicating a general improvement in risk score over the course of the surveys) and one representing negative changers (those with a negative slope, indicating a worsening of risk score).

Predictor Variables

Predictor variables used in our analysis include six baseline sociodemographic variables (age, education, marital status, ethnicity, sex, city of residence) and five baseline psychosocial variables (cardiovascular disease health knowledge, health attitudes, self-efficacy, perceived risk, and health media use).

Information on sociodemographic variables and psychosocial variables was obtained from self-reported questionnaire data. Age and education were collected as continuous variables and marital status (currently married or not currently married) and ethnicity (White, Hispanic) were collected as dichotomous variables. The 2% of the sample who were Asian and the 2% who were African American were excluded from this analysis. Education was selected as the measure of socioeconomic status because education is available for all individuals regardless of employment status, has high reliability and validity, is generally stable after early adulthood,[25,26] and has recently been shown to have a stronger association with cardiovascular disease risk factors than income or occupation.[27,28]

Nine psychosocial predictor variables were initially selected for analyses—these were reduced to five by means of a principal components analysis to reduce collinearity. Health knowledge was measured on a summative scale of 17 items about cardiovascular disease risk factors (Cronbach's $\alpha = .66$). Health attitudes were measured on a summative scale encompassing four dimensions—8 physical activity items, 7 weight items, 10 nutrition items, and 10 stress management items ($\alpha = .83$). Self-efficacy was based on 13 diet and exercise items ($\alpha = .79$), and perceived cardiovascular disease risk was measured by one item (scale of 1 to 7). Health media use, which assesses the frequency with which health information is obtained from newspapers, magazines, and pamphlets sent in the mail, was based on four items and was standardized to a mean of 0 and a standard deviation of 1 ($\alpha = .70$). Higher scores reflect higher health knowledge, more positive health attitudes, higher self-efficacy, greater perceived risk, and higher health media use.

Statistical Model

A multiple logistic regression model including interaction terms is a common analytic method for identifying characteristics of respondents who show favorable risk factor changes compared with those who do not. We used this method initially, but we found many significant higher-order interaction terms, making results difficult to interpret. In addition, ordering individual subjects on a scale, as the risk score does, is not as useful for planning and implementing prevention programs as is the identification of specific subgroups. To overcome these problems, we used a signal detection model that provides more interpretable results when there are a large number of interactions.[7] This method, which is a form of recursive partitioning, uses baseline characteristics to define distinct subgroups of respondents, which are mutually exclusive and maximally discriminated from each other, on the basis of probability of positive change in risk function score.

The baseline sociodemographic and psychosocial predictor variables were entered into the signal detection analysis along with minimum and maximum values and interval cutpoints (further information is available from the authors upon request). The signal detection algorithm

then examined each variable and its possible cutpoints and selected a variable and cutpoint on the basis of a combined optimal measure of sensitivity and specificity with regard to the outcome measure (cardiovascular disease risk factor change).

After choosing and splitting on the first optimally efficient variable, the signal detection program separately searched each subgroup or "branch" of the first split for the next most efficient variable and cutpoint (Figure 1). This procedure was repeated separately in each subgroup with all the remaining predictor variables. The procedure ended when (1) there were too few subjects in a subgroup for further analysis, (2) no further significant discriminating variable at ($P < .05$) could be found, or (3) no further predictor variables remained.

Results

The cohort treatment sample was composed of 221 women and 190 men aged 25 through 74 years who participated in at least the baseline and final surveys. Of those who participated in the first survey, 70.0% returned for the second survey, 58.7% for the third survey, and 54.0% for the final survey.

The baseline sample consisted predominantly of White adults (92%) who had completed at least high school (86%) and who were married (71%). Sixty-nine percent (n = 284) of the study population showed a positive change in their composite risk factor scores over the 6-year study period.

Figure 1 shows the development of the optimally efficient algorithm for identifying distinct groups on the basis of positive change in cardiovascular disease risk factor score. The first optimally efficient variable ($P < .001$) that distinguished positive changers from negative changers was age (≤55 years vs >55 years). In the older age group (group 1), 83% of the participants showed positive change. No other variable provided significant discrimination within this group.

Among younger participants (≤55 years of age), educational attainment further discriminated respondents. For those with higher educational attainment (>11 years), the next significant split was found for self-efficacy. This variable subdivided the younger subgroup with higher educational attainment into two smaller groups. In the group with lower self-efficacy scores (group 2), 70% showed positive change; in the group with higher scores (group 3), 55% showed positive change. The last subgroup of younger participants was discriminated on the basis of lower educational attainment (≤11 years). This group (group 4) had the lowest proportion of positive changers (42%). After the identification of these four subgroups, no further significant predictors were found and the signal detection test development stopped.

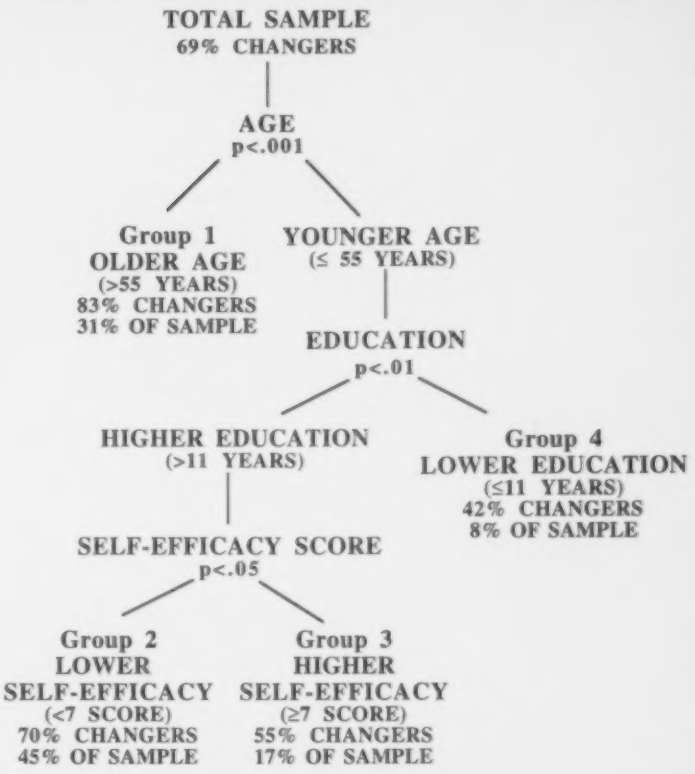


FIGURE 1—Signal detection analysis: development of the optimally efficient algorithm for identifying distinct groups on the basis of positive change in cardiovascular disease risk factor score.

As in all correlational analyses, the factors selected identify respondents at different risk levels but are not necessarily themselves causal factors. Thus, to gain further understanding of the nature of the four subgroups, we examined the profile of each (Table 1). Group 1, the oldest age group and the most likely to show a positive overall risk function change, was distinguished from the other subgroups by higher cholesterol and blood pressure levels and higher prevalence of hypertension at baseline. Group 1 also had the highest perceived risk and highest health media use scores. Group 2, which had lower self-efficacy scores, had a higher proportion of Hispanics, a higher proportion of smokers and heavy smokers, higher rates of hypertension, and higher levels of cholesterol than did group 3. Members of group 3, which had the highest self-efficacy scores, were the youngest participants, the most highly educated, the most likely to be single, and the most likely to be male. They also had the lowest perceived risk and the most positive risk factor profile. Group 4, the group with the lowest proportion of positive changers, had the highest proportion of women, the highest proportion of Hispanics, and the lowest levels of health knowledge.

Discussion

In this paper we provide a prospective examination of factors associated with changes in overall cardiovascular disease risk factor scores in individuals living in

TABLE 1—Baseline Characteristics of Four Change Groups: Men and Women Aged 25 through 74 Years, Stanford Five-City Project, 1979 through 1985

	Group 1: Older Adults	Group 2: Younger, More Educated, Lower Self-Efficacy	Group 3: Younger, More Educated, Higher Self-Efficacy	Group 4: Younger, Less Educated	p[a]
No. in group	127	184	69	31	
Positive changers, %	83	70	55	42	
Sociodemographic variables					
Mean age, y (SD)	63.3 (5.0)[b]	39.4 (9.3)[b]	36.0 (8.1)[b]	40.4 (10.0)[b]	.001
Mean education, y (SD)	13.1 (3.6)	14.3 (2.3)[b]	15.1 (2.4)[b]	8.1 (2.6)[b]	.001
Married, %	70.1	75.5	53.6	80.6	.004
Hispanic, %	4.7	7.6	1.5	38.7	.001
Women, %	55.1	55.4	40.6	67.7	.06
Physiologic variables					
Smokers, %	22.2	32.1	26.1	32.3	.27
Mean no. cigarettes/d (smokers) (SD)	16.5 (10.2)	22.2 (12.9)	12.2 (8.5)	17.7 (9.8)	.02
Mean diastolic blood pressure, mm Hg (SD)	81.9 (11.1)	77.6 (10.0)	76.8 (10.7)	77.5 (10.8)	.01
Hypertensive, %	57.6	23.9	15.9	23.3	.001
Mean plasma cholesterol, mg/dL (SD)	224.5 (37.8)	200.3 (38.7)	192.9 (38.3)	199.7 (41.7)	.001
Mean pulse, beats/min (SD)	69.9 (12.3)	69.7 (9.6)	69.0 (8.7)	70.2 (8.8)	.94
Psychosocial variables					
Mean cardiovascular disease knowledge score (SD)	6.9 (3.0)	6.9 (2.7)	6.9 (2.9)	4.6 (2.3)	.001
Mean health attitudes score (SD)	4.2 (0.7)	3.8 (0.6)	4.3 (0.7)	3.7 (0.7)	.001
Mean self-efficacy score (SD)	6.0 (1.5)	5.4 (1.1)[b]	7.8 (0.6)[b]	4.8 (1.5)	.001
Mean perceived risk score (SD)	4.2 (1.6)	3.7 (1.3)	3.3 (1.4)	3.9 (1.4)	.001
Mean health media use score (SD)[c]	0.2 (0.7)	−0.2 (0.7)	−0.1 (0.8)	−0.1 (0.8)	.001

[a]Based on analysis of variance for continuous variables and chi-square tests for categorical variables.
[b]Numbers that segmented the sample into subgroups in the signal detection model.
[c]Standardized to a mean of 0 and a standard deviation of 1.

the two Stanford Five-City Project treatment cities that received the 6-year education program. Our objective is to provide researchers who are planning cardiovascular disease community interventions with information about subgroups that are the most likely (or unlikely) to make positive changes in response to intervention programs, especially those involving media interventions. Although other studies have examined the effectiveness of demographic and psychosocial variables in segmenting populations,[29,30] few have examined the relationship between segmentation variables and risk factor changes over time.

Four distinct subgroups of individuals were identified on the basis of changes in their cardiovascular disease risk factor scores. Group 1 (31% of the sample) consisted of older adults who initially had the highest mean blood pressure levels, double the rate of hypertensives of other groups, and the highest mean cholesterol levels. This group had the most positive motivation and information-seeking habits, as indicated by its high perceived risk and health media use scores. Given their psychosocial and physiologic profiles, it is not surprising that members of this group were the most likely to change; therefore, some portion of the change may reflect a regression to the mean phenomenon. Their perception of risk was accurately high, their cardiovascular disease knowledge was adequate, and their health media use was high (providing a relatively information-rich environment via print materials). Moreover, their self-efficacy scores show a confidence about making behavioral changes needed to lower risk. The demographic and psychosocial profile of members of this subgroup, combined with their health media use, make them receptive to public health campaigns that disseminate risk reduction information.

Members of groups 2 and 3, who constituted 62% of the sample, were identified by younger age and higher educational attainment. Group 2 (with 70% changers) was distinguished from group 3 (with 55% changers) by self-efficacy score (5.4 and 7.8, respectively). Group 2 also had a higher smoking rate (32%, vs 26% in group 3), a higher percentage of hypertensives (24%, vs 16% in group 3), and a higher mean cholesterol level (200.3 mg/dL in group 2 and 192.9 mg/dL in group 3). The higher percentage of positive changers in group 2 than in group 3 is most likely explained by group 2's greater cardiovascular disease risk at baseline, which gives a greater possibility for change.

At first glance, the fact that group 3 had a higher level of self-efficacy and a lower proportion of positive changers seems counter to the large empirical literature on self-efficacy and behavior change.[9,14] On closer examination, however, we see that group 3 was the youngest of all groups, the most educated, and the most likely to be single, non-Hispanic, and male. This group was also the healthiest in terms of blood pressure and cholesterol. Thus, while its members' self-efficacy and health attitudes were highest, their low

levels of perceived risk may indicate that they believe they could change but do not see a need to do so. It is also possible that members of this group have already made changes and do not see the need for further change. The effort to reach those who are younger, more educated, and healthier, but less likely to seek health information via the media, may be most successful if prevention is stressed via mainstream formats such as newspaper articles rather than specialized health media such as doctors' columns. Furthermore, strategies that incorporate short-term outcomes and use a social influences framework may be most appealing to this audience.

Like group 1, group 4 is one that could be labeled "high risk," but unlike group 1, group 4 had a low rate of change (42%). Its members were the least educated of all groups (mean of 8.1 years of schooling). Although predominantly White, almost 40% were Hispanic. Over 65% were women. Members of this group had the highest smoking rates, the lowest levels of cardiovascular disease knowledge, the lowest health attitude scores, and the lowest self-efficacy scores. Although it constitutes only 8% of the sample, this group is at high risk for health problems because of its lower socioeconomic status[27,31] and poorer health knowledge and attitudes. Although persons of all educational levels are beginning to show substantial declines in cardiovascular disease risk factors,[32] those with low educational attainment may be unlikely to respond to health education campaigns that are based on a high level of cardiovascular disease knowledge and the initiation of complex actions. It is possible that members of this subgroup were not provided with the requisite skills to perform many of the behaviors promoted by the campaign and that they lived in environments where resources were scarce and where norms did not support positive changes in health. It is clear that more research is needed on how to provide effective health education to those who fit the profile of this subgroup. For example, community interventionists need to gain a better understanding of how poverty and literacy levels, normative beliefs, and communication styles affect behavior change. Furthermore, interventionists need to increase their understanding of the social environments in which adults from lower educational levels reside.[33-37]

Strengths and Limitations

A major strength of the Five-City Project is the opportunity it offers to conduct cohort analyses on large numbers of men and women for whom multiple sociodemographic, psychosocial, and physiologic measurements are available. In addition, participants in the Project are more likely to represent the community than are clinical or volunteer samples, thus enhancing the study's generalizability. Because of the study's location and the type of intervention, the results are most relevant to California communities with campaigns similar to the Five-City Project.

As noted, only 54.0% of those who participated in the baseline survey (1979/80) also participated in the final cohort survey (1984/85). Although the majority of dropouts were out-migrants rather than refusers, dropouts were significantly different from nondropouts on a number of sociodemographic variables. For example, dropouts were significantly younger (mean age 42.2 vs 46.6 years) and less educated (mean years of education 12.5 vs 13.4) than nondropouts.

Another potential limitation of the study is the possible bias that could arise if those with more motivation to change their cardiovascular disease risk factors were more likely to continue participating in the study. We assessed this potential source of error by comparing respondents who participated in at least the baseline and final cohort surveys with those who dropped out after the baseline survey. We found no significant differences between dropouts and nondropouts on a summary index of nine items assessing confidence in changing dietary habits ($P = .15$) or on a single item assessing intention to quit smoking ($P = .71$, smokers only).

Public Health Implications

This study has three important public health intervention implications. First, this research reinforces the work of others in pointing out the need for campaign designers to segment audiences into more homogeneous and identifiable subgroups to enhance the likelihood that each subgroup will be reached effectively. Second, it points out the need to use sociodemographic, psychosocial, and physiologic variables when segmenting audiences. Third, the distinct makeup of the four population subgroups illustrates that communitywide health education campaigns need to develop specific interventions that target different age, socioeconomic, and cultural subgroups to enhance the likelihood that all will be reached. □

Acknowledgments

This research was supported by Public Health Service grant 1RO1-HL-21906 from the National Heart, Lung, and Blood Institute to Dr John W. Farquhar.

The authors gratefully acknowledge the assistance of Drs Stephen P. Fortmann and Erica Frank for valuable comments on the manuscript; Beverly Rockhill, Darius Jatulis, and Jennifer Lin for statistical advice; Patricia Cruz for the preparation of tables; and Donna Balderston, Christine Brereton, Doreen Chapman, Mary Collins, Kathryn Kandler, Jennefer Santee, Sarjeet Singh, and Sammy Van Gundy for their valuable participation in all phases of recruitment and data collection.

References

1. Farquhar JW, Fortmann SP, Maccoby N, et al. The Stanford Five-City Project: design and methods. *Am J Epidemiol.* 1985;122:323–334.
2. Farquhar JW, Fortmann SP, Flora JA, et al. Effects of communitywide education on cardiovascular disease risk factors: the Stanford Five-City Project. *JAMA.* 1990; 264:359–365.
3. Fortmann SP, Winkleby MA, Flora JA, Haskell WL, Taylor CB. Effect of long-term community health education on blood pressure and hypertension control: the Stanford Five-City Project. *Am J Epidemiol.* 1990;132:629–646.
4. Fortmann SP, Taylor CB, Flora JA, Jatulis DE. Changes in adult cigarette smoking prevalence after 5 years of community health education: the Stanford Five-City Project. *Am J Epidemiol.* 1993;137:82–96.
5. Blackburn H, Luepker RV, Kline FG, et al. The Minnesota Heart Health Program: a research and demonstration project in cardiovascular disease prevention. In: Matarazzo JD, Miller NE, Weiss SM, eds. *Behavioral Health: A Handbook of Health Enhancement and Disease Prevention.* New York, NY: John Wiley & Sons; 1984:1171–1178.
6. Carleton RA, Lasater TM, Assaf AR, Lefebvre RC, McKinlay SM. The Pawtucket Heart Health Program: an experiment in population-based disease prevention. *R I Med J.* 1987;70:533–538.
7. Kraemer HC. *Evaluating Medical Tests: Objective and Quantitative Guidelines.* Newbury Park, Calif: Sage Publications; 1992.
8. Bandura A. Self-efficacy mechanisms in human agency. *Am Psychol.* 1982;37:122–147.
9. Bandura A. *Social Foundations of Thought and Action.* Englewood Cliffs, NJ: Prentice-Hall; 1986.
10. McGuire W. Attitudes and attitude change. In: Lindzey G, Aronson E, eds. *The Handbook of Social Psychology.* 3rd ed. Hillsdale, NJ: Lawrence Erlbaum; 1985.
11. Lefebvre RC, Flora JA. Social marketing and public health intervention. *Health Educ Q.* 1988;15:299–315.
12. Fishbein M, Ajzen I. *Belief, Attitude, Intention and Behavior.* Menlo Park, Calif: Addison Wesley; 1975.

13. Roser C, Flora JA, Chaffee SH, Farquhar JW. Using research to predict learning from a PR campaign. *Public Relations Rev.* 1990;Summer:61–77.
14. Maibach E, Flora JA, Nass C. Changes in self-efficacy and health behavior in response to a minimal contact community health campaign. *Health Commun.* 1991;3: 1–15.
15. Chaffee SH, Roser C. Involvement and the consistency of knowledge, attitudes and behaviors. *Commun Res.* 1986;13:373–399.
16. Jatulis D, Flora JA, Winkleby MA, Fortmann SP. Who remembers heart disease programs? Presented at the 119th Annual Meeting of the American Public Health Association; November 12, 1991; Atlanta, Ga.
17. Fortmann SP, Marcuson R, Bitter PH, Haskell WI. A comparison of the Sphygmetrics SR-2 automatic blood pressure recorder to the mercury sphygmomanometer in population studies. *Am J Epidemiol.* 1981;114:836–844.
18. Fortmann SP, Rogers T, Haskell WL, Solomon DS, Vranizan K, Farquhar JW. Indirect measures of cigarette use: expired air carbon monoxide vs. plasma thiocyanate. *Prev Med.* 1984:13;127–135.
19. Flora JA, Maccoby N, Farquhar JW. Communication campaigns to prevent cardiovascular disease: the Stanford community studies. In: Rice RE, Atkin CK, eds. *Public Communication Campaigns.* Newbury Park, Calif: Sage Publications; 1989: 233–252.
20. Chaffee SH, Roser C, Flora JA, Farquhar JW. A functional interpretation of a health campaign's effects. *Health Commun.* In press.
21. Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. *J Chron Dis.* 1967:20;511–524.
22. Winkleby MA, Fortmann SP, Rockhill B. Cigarette smoking trends in adolescents and young adults: the Stanford Five-City Project. *Prev Med.* 1993;22:325–334.
23. *Lipid Research Clinics Manual of Laboratory Operations. Vol 1. Lipid and Lipoprotein Analysis.* Washington, DC: US Dept of Health, Education, and Welfare; 1974.
24. Gibbons RD, Hedeker D, Waternaux C, et al. Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Arch Gen Psychiatry.* 1993;50:739–750.
25. Mueller CW, Parcel TL. Measures of socioeconomic status: alternatives and recommendations. *Child Dev.* 1981;52:13–30.
26. Liberatos P, Link BG, Kelsey JL. The measurement of social class in epidemiology. *Epidemiol Rev.* 1988;10:87–121.
27. Winkleby MA, Fortmann SP, Barrett DC. Social class disparities in risk factors for disease: eight-year prevalence patterns by level of education. *Prev Med.* 1990;19:1–12.
28. Winkleby MA, Jatulis DE, Frank E, Fortmann SP. Socioeconomic status and health: how education, income, and occupation contribute to cardiovascular disease. *Am J Public Health.* 1992;82:816–820.
29. Slater MD, Flora JA. Health lifestyles: audience segmentation analysis for public health interventions. *Health Educ Q.* 1991; Summer:221–233.
30. Snyder LB, Rouse RA. Target the audience for AIDS messages by actual and perceived risk. *AIDS Educ Prev.* 1992;4: 160–172.
31. Marmot MG, McDowall ME. Mortality decline and widening social inequalities. *Lancet.* 1986;ii:274–276.
32. Winkleby MA, Fortmann SP, Rockhill B. Trends in cardiovascular disease risk factors by educational level: the Stanford Five-City Project. *Prev Med.* 1992;21:592–601.
33. Rogers EM, Shoemaker FF. *Communication of Innovations: A Cross-Cultural Approach.* 2nd ed. New York, NY: MacMillan; 1971.
34. Wallack L, Winkleby MA. Primary prevention: a new look at basic concepts. *Soc Sci Med.* 1987;25:923–930.
35. Bunker JP, Gomby DS. Preface. In: Bunker JP, Gomby DS, Kehrer BH, eds. *Pathways to Health: The Role of Social Factors.* Menlo Park, Calif: Henry J. Kaiser Foundation; 1989:15.
36. Dervin B. Audience as listener and learner, teacher and confidante: the sense-making approach. In: Rice RE, Atkin CK, eds. *Public Communication Campaigns.* Newbury Park, Calif: Sage Publications; 1989: 67–86.
37. Freimuth VS, Mettger W. Is there a hard-to-reach audience? *Public Health Rep.* 1990;105:232–238.

Effects of a Restricted Work-Site Smoking Policy on Employees Who Smoke

Janet Brigham, PhD, Janet Gross, PhD, Maxine L. Stitzer, PhD, and Linda J. Felch, MA

ABSTRACT

Objectives. This study evaluated the biological and subjective consequences observed in individual smokers after implementation of a workplace smoking-restriction policy.

Methods. Employees were evaluated for 4 weeks before and 4 weeks after their workplace became smoke-free (n = 34). A comparison group of smokers whose work-site smoking was unrestricted served as controls (n = 33). Daily exposure to tobacco constituents and withdrawal effects were measured.

Results. Smokers at the restricted site had verified smoking reduction (mean = four cigarettes per day) and significantly reduced nicotine and carbon monoxide during the work shift. There were increases in ratings of some common withdrawal symptoms (cravings/urges, concentration difficulties, increased eating, depression). No evidence of compensatory smoking during nonwork hours was found. Overall tobacco exposure, as measured in saliva cotinine, showed a nonsignificant 15% decline.

Conclusions. Workplace smoking restriction markedly altered smoking patterns (i.e., reduced daytime smoking) and reduced cotinine levels to an amount consistent with cigarette reduction. Thus, work-site smoking restriction may promote meaningful, albeit limited, reductions in tobacco exposure and consequent health risks. (*Am J Public Health.* 1994;84:773–778)

Introduction

Many institutions throughout the United States have moved toward restricting smoking in public places, particularly since the Surgeon General's 1986 report on involuntary smoking.[1] Several studies have indicated that the incidence of public smoking changes when restricted-smoking policies are implemented. Stillman et al.[2] reviewed empirical evidence of reduction in smoking at the Johns Hopkins medical institutions and concluded that implementing a smoke-free policy in a large medical center can decrease visible smoking and passive exposure to tobacco smoke. A previous report by Becker et al.[3] indicated that public smoking at the Johns Hopkins Children's Center was virtually eliminated in 1987 by implementation of a smoke-free policy. Recently, Borland et al.[4] reported that Australian smokers exposed to a mandatory work-site smoking ban reduced their smoking by about seven cigarettes per day while working inside and by about five cigarettes per day overall with very little compensatory, or increased, smoking observed outside of work after enforced reduction at work.

Of interest to policymakers and work-site managers are the potential negative side effects of daytime abstinence, which may affect the smoker's behavior (e.g., symptoms of the tobacco withdrawal syndrome), as well as the overall impact on smoking exposure (e.g., compensatory changes in rates of smoking outside the work place). Researchers on smoking cessation have found that some of the symptoms of the tobacco withdrawal syndrome that typically accompany tobacco abstinence[5] are present to a lesser extent in individuals who reduce cigarette consumption or switch to a lower-nicotine-yielding cigarette.[6,7] For example, in the study by Hatsukami et al.,[6] smokers who reduced the number of cigarettes smoked by 50% or reduced the nicotine yield of their cigarettes by about 50% experienced some craving and withdrawal discomfort but reported significantly less discomfort than total abstainers. On the other hand, West et al.[7] found that switching to an ultra-low-yield cigarette did not result in any appreciable reporting of withdrawal distress, except for increased hunger. The mechanisms thought to cause low-level withdrawal symptoms under conditions of restricted smoking are reduction in blood nicotine levels and habit change.

In a related vein, research on restricted smoking in habitual smokers has suggested that smokers may adjust their smoking behavior to offset effects of restricted access or reduced cigarette yield.[8,9] Benowitz et al.[10] found that when smokers reduced their consumption from an average of 37 to 5 cigarettes per day (i.e., a sixfold reduction), the toxicity per cigarette increased threefold because of more puffs per cigarette, longer breath holding, and deeper inhalation, whereas the cumulative daily exposure to tar and carbon monoxide dropped by only about one half. This tendency for smokers to alter their smoking style (e.g., by increasing inhalation or taking more frequent

Janet Gross, Maxine L. Stitzer, and Linda J. Felch are with the Johns Hopkins University School of Medicine, Baltimore, Md. At the time of this study, Janet Brigham was with the Johns Hopkins University School of Medicine; currently she is with the Western Psychiatric Institute, University of Pittsburgh, Pittsburgh, Pa.

Requests for reprints should be sent to Maxine L. Stitzer, PhD, Behavioral Pharmacology Research Unit, 5510 Nathan Shock Dr, Baltimore, MD 21224.

This paper was accepted August 3, 1993.

TABLE 1—Baseline Characteristics of Groups (n = 67)

	Group	
	Restricted (n = 34)	Control (n = 33)
Mean age, y (SD)	38 (9.5)	37 (10.4)
Mean weight, lb (SD)	176 (43)	173 (48)
Occupational status,[a] %		
1	12	10
2	30	29
3	42	26
4	15	35
Gender, % male	27	30
Mean scores on self-report (SD)		
Tension-Anxiety	7 (6.2)	10 (8.0)
Depression-Dejection*	5 (6.6)	10 (11.0)
Anxiety-Hostility	7 (8.6)	7 (7.7)
Vigor	17 (5.4)	16 (7.8)
Fatigue	8 (6.6)	8 (6.7)
Confusion-Bewilderment	4 (4.4)	5 (4.8)
Friendliness	19 (3.7)	19 (4.7)
Total Mood Disturbance	13 (31.8)	25 (39.3)
Total withdrawal score	11 (7.1)	15 (9.6)
Smoking characteristics, mean (SD)		
Years smoked	19 (8.6)	18 (10.4)
No. of cigarettes/day*	22 (10.7)	27 (10.0)
Carbon monoxide, ppm	26 (14.0)	29 (12.6)
Saliva nicotine, ng/mL	566 (764)	595 (640)
Saliva cotinine, ng/mL	232 (136)	282 (121)

[a]1 = low status and 4 = high status.
*$P < .05$.

puffs) could offset the potential health advantages of reducing the number of cigarettes smoked at work. The extent to which smokers engage in compensatory smoking behavior and increase their exposure to toxic constituents of tobacco as a result of restrictions on workplace smoking is unknown.

Among the hospitals joining the smoke-free movement recently was Francis Scott Key Medical Center, part of the Johns Hopkins medical institutions in Baltimore, Md. This institution implemented a hospital-wide smoking ban on July 1, 1989. We used this event as an opportunity to examine the biological, subjective, and behavioral impact of a smoking policy restriction on individual smokers, including objective measures of tobacco smoke exposure. We compared these effects with those in a control group of smokers whose work-site smoking policy remained unrestricted.

Methods

Subjects

Ninety-two smokers were recruited for the present study. All study participants were between 18 and 65 years old and were not seeking treatment to quit smoking. Forty-one smokers were recruited from the Francis Scott Key Medical Center, the site of the smoking ban, through advertisements, flyers, and referrals. Before analysis, four dropped out of the study, and three were omitted from the analysis because of incomplete compliance with the study procedures. The comparison control group was composed of 51 individuals recruited through newspaper advertisements at several nearby hospitals that had no restricted smoking policy. Of those initially enrolled, 10 failed to complete the study, 1 quit smoking during the study, and 7 were excluded from analyses because their mean number of cigarettes per day before the smoking ban exceeded the maximum number observed in the restricted-smoking group. Consequently, 67 participants were included in the final analysis of the data.

Table 1 shows the demographic, smoking, and psychological characteristics of the two groups at baseline. The groups were similar with respect to mean age, distribution of gender, and occupational status. The self-reported amount of smoking was significantly higher in the control group (t (60) = 2.02, $P < .05$). However, the groups did not differ on the biological indices of tobacco exposure, which included expired-breath carbon monoxide and saliva cotinine and nicotine. The groups also differed with respect to the total withdrawal score and the score on the Depression-Dejection subscale of the Profile of Mood States. In both cases, the control subjects had higher scores, indicative of greater distress.

Procedures

All subjects participated in data collection sessions at the research laboratory after their work shifts. Experimental (restricted smoking) subjects attended a laboratory data collection session once per week for 4 weeks before and 4 weeks after implementation of the smoking ban. Control subjects attended sessions at their work site once weekly for 8 consecutive weeks. Informed consent, which had been approved by the Francis Scott Key Medical Center Institutional Review Board, was obtained from all subjects after the study was explained at the first session. Subjects were paid $100 for successfully completing the study.

Measures

At each study contact, behavioral, biological, and subjective report data were collected.

Smoking amount. Smoking amount was assessed via self-report and by collection of cigarette butts for designated time periods. To determine the pattern of smoking during the work shift, subjects were asked at each lab visit to recollect and report the number of cigarettes smoked at their workstation, at other sites inside the building, and at any sites outside the building for the time period of the most recent work shift. To estimate 24-hour smoking amount and toxic exposure to tobacco constituents, subjects collected their cigarette butts for the 24-hour period before the lab session. Three separate plastic collection bags reflected three intervals of the day: (1) between the end of work and bedtime the previous day, (2) between arising and going to work the day of the lab session, and (3) during work the day of the lab session. The number of cigarette butts per interval and the weight of the butts from each interval were assessed. Thus, for the period of the work shift, both self-report and butt count data were available. Butt

weight was used as an indirect measure of smoking intensity.[9] Like butt length, butt weight indicates the amount of material in the cigarette left unsmoked. Butt length provides an indirect measure of smoking intensity and relative toxicity because tar and carbon monoxide delivery increase logarithmically as cigarettes are smoked to shorter butt lengths.[11] Before weighing, all butts were desiccated until the moisture content in the desiccator was below 20%.

Biological exposure. At each lab session, subjects provided saliva samples, which were then analyzed for nicotine and cotinine (a metabolite of nicotine with a half-life of 18–20 hours) by using gas chromatography (analyses conducted by Labstat, Inc, Kitchener, Ontario, Canada). Breath carbon monoxide samples, which reflected recent exposure to tobacco smoking, were measured by using a Vitalograph breath analyzer (Lenexa, Kans).

Subjective report. Three subjective measures were used to assess withdrawal symptoms associated with smoking reduction and the impact of those symptoms on participants' work. A withdrawal symptom questionnaire adapted from that used by Hughes and Hatsukami[5] assessed subjects' ratings of 18 symptoms (0 = none to 3 = severe), many of which characterize the DSM-III Tobacco Withdrawal Syndrome. This measure has been widely used in smoking research and has good psychometric properties.[12] The Profile of Mood States,[13] a widely used 65-item self-report questionnaire, was used to measure overall mood disturbance. This questionnaire yields factor scores on subscales of Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigor, Fatigue, Confusion-Bewilderment, Friendliness, and Total Mood Disturbance. A work attitude and performance questionnaire developed by the investigators solely for purposes of this study was designed to elicit subjects' ratings (on a scale of 0 to 10) regarding their concentration ability and productivity at work, relations with coworkers, work enjoyment and satisfaction, and overall sense of well-being.

Statistical Analysis

We analyzed the data with a two-factor, mixed-model analysis of variance using SPSS-X for the Macintosh. The grouping factor was restricted smoking (smoking ban) vs control (unrestricted smoking); the repeated-measures, within-subject factor was time: preban (weeks 1–4) vs postban (weeks 5–8). All tests of significance were two tailed. Posthoc comparisons of preban and postban interactions were conducted by using Tukey's Honestly Significant Difference (HSD) test.

Results

Work-Site Smoking

Amount and intensity. Based on the self-reported smoking amount, consumption of cigarettes during work hours by subjects in the restricted group dropped significantly as a function of the ban ($F[1, 65] = 15.7$, $P < .0001$). Restricted subjects reported an average of 7.57 (SD = 4.7) cigarettes per day during study weeks 1–4 (before the ban) vs 3.64 (SD = 4.9) cigarettes per day during study weeks 5–8 (after the ban) (Tukey's HSD $q = 9.15$, $P < .001$). None of the subjects in the restricted group quit smoking during the study. The smoking pattern for control subjects did not differ over time. The mean rate of smoking was 10.02 (SD = 2.9) cigarettes per day before the ban vs 9.53 (SD = 4.8) cigarettes per day after the ban. Reported smoking rates were highly similar to those determined with the butt count procedure for this time interval. Although there was a small discrepancy between self-report and butt count with respect to absolute numbers of cigarettes, there was in both cases a four-cigarette-per-day reduction in smoking during work hours in the restricted group as a function of the ban, vs less than a one-cigarette-per-day change in the control group.

The average butt weight per cigarette smoked during work was calculated from the returned butts for this period. There were no effects on butt weight found from the repeated-measures analysis of variance (ANOVA). In the restricted group, the mean butt weight was 0.36 g (SD = 0.11 g) before the ban and 0.35 g (SD = 0.09 g) after the ban. For the control group, the means were 0.38 g (SD = 0.06 g) and 0.36 g (SD = 0.08 g), respectively.

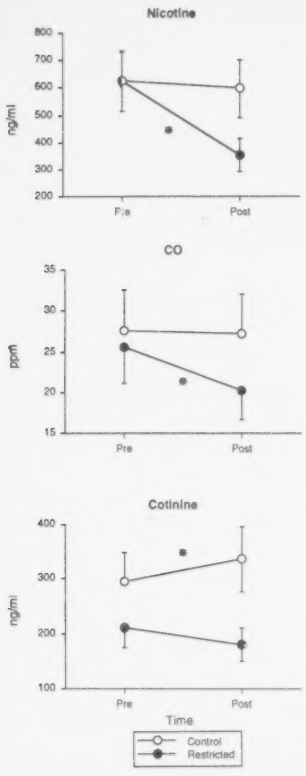
Smoking location. An analysis of self-reported amount of smoking at the various work-shift locations showed significant interactions for smoking at the workstation ($F[1, 65] = 41.6$, $P < .0001$) and for smoking outside the building ($F[1, 65] = 19.6$, $P < .0001$). Restricted subjects reported an average of 5.89 (SD = 4.8) cigarettes per day at the workstation before the ban compared with 0.69 (SD = 2.2) cigarettes after the ban ($q = 12.25$, $P < .01$). The reverse pattern was observed for their smoking behavior outside the building. Before the ban, the restricted group smoked an average of 0.83 (SD = 1.0) cigarettes per day outside the building compared with 2.51 (SD = 1.7) cigarettes after the ban ($q = 6.7$, $P < .01$). The ban on smoking had no effect on behavior at other locations within the building. Restricted subjects reported a mean of 0.84 (SD = 1.1) cigarettes per day before the ban and a mean of 0.45 (SD = 1.2) cigarettes per day after the ban at other building sites. Control subjects did not change their smoking patterns over time (mean number of cigarettes at workstation = 6.8; mean at other inside locations = 1.2; mean outside building = 1.8).

Before implementation of the smoking ban, 82% of the subjects in the restricted group reported ever smoking at their workstations. After implementation of the ban, this figure decreased to 18%. The percentage of control subjects smoking at the workstation during these two time periods was 88% and 85%, respectively. The percentage of subjects who reported going outside the building to smoke during the workday increased in the restricted group from 68% to 97% as a function of the ban. The percentage of control subjects reporting smoking outside the building during these two time periods changed only slightly from 55% to 49%.

Biological exposure. The repeated-measures ANOVA revealed significant interactions on both saliva nicotine and breath carbon monoxide measures, which reflected changes in work-shift smoking behavior. As shown in Figure 1, average saliva nicotine measured at the end of the work shift in the restricted group decreased from 623 ng/mL before the ban to 354 ng/mL after the ban ($q = 5.11$, $P < .01$), whereas there was no change in the control group. There was also a significant decrease in breath carbon monoxide in the restricted group from 25.6 to 20.2 ppm ($q = 5.43$, $P < .01$) vs no change in the control group.

Smoking during Off-Work Hours

Amount and intensity. Butt count data revealed that the number of cigarettes smoked per day during off-work hours did not change significantly in either group as a function of the ban (mean number of cigarettes [restricted group] = 12.56 preban [SD = 5.7] vs 11.44 postban [SD = 6.4]; mean number of cigarettes [control group] = 11.99 preban [SD = 4.4] vs 11.14 postban [SD = 4.8]). Nor did the mean butt weight of cigarettes collected



Note. Bars reflect the standard deviation. The preban point represents the average of four weekly data collections during weeks 1–4; the postban point represents the average of four weekly data collections during weeks 5–8. Participants in the restricted group were banned from smoking at the work site during the postban period. Asterisks indicate differences between the means according to Tukey's HSD. Both nicotine and CO levels changed significantly across the time periods in the restricted group ($P < .01$).

FIGURE 1—Saliva nicotine, expired breath carbon monoxide (CO), and saliva cotinine levels are shown for the preban and postban time periods.

from the before-work and after-work assessment periods differ as a function of the smoking ban implementation. For the restricted group, the mean butt weight per cigarette for cigarettes smoked outside work before the ban was 0.33 g (SD = 0.11 g) vs 0.34 g (SD = 0.13 g) for the period after the ban. For the control group, the mean outside-of-work butt weights before and after the ban were 0.36 g (SD = 0.12 g) and 0.33 g (SD = 0.14 g), respectively. The mean weight of butts collected after work was highly similar to the mean weight of butts collected before work.

Biological exposure. Cotinine levels were significantly higher in the control group overall ($F[1, 65] = 13.7, P < .0001$), and there was a significant interaction between group and time ($F[1, 65] = 17.0, P < .0001$). As shown in Figure 1, cotinine, which reflects overall nicotine exposure for the previous several days, decreased by 15% in the restricted group from 211 to 180 ng/mL ($q = 3.61, P > .05$) and increased by 13.5% in the control group from 296 to 336 ng/mL ($q = 4.62, P < .01$). Subjects were categorized according to whether their average cotinine levels increased or decreased from the preban to the postban period. In the restricted group, 71% of the subjects showed a reduction in cotinine levels, with a mean decrease of 55.9 ng/mL (range = 0.7–120.6 ng/mL). In the control group, 33% of the subjects showed a reduction in cotinine levels, with a mean decrease of 39.8 ng/mL (range = 8.1–137.7 ng/mL). In the restricted group, 29% of the subjects showed an increase in cotinine levels, with a mean increase of 28.4 ng/mL (range = 7.4–83.6 ng/mL). In the control group, 67% of the subjects showed an increase in cotinine levels, with a mean increase of 80.5 ng/mL (range = 1.3–227 ng/mL). A chi-square test on the number of subjects with increasing vs decreasing cotinine levels was significant (χ^2 [1] = 9.32, $P < .002$).

Subjective Reports

Withdrawal symptom questionnaire. Table 2 shows the mean values and standard deviations for the withdrawal items with significant interaction terms, significant changes from the preban to the postban period in the restricted group, and significant group differences during the postban period. Restricted subjects reported significant increases in 4 of 18 withdrawal symptom items (craving for cigarettes, difficulty concentrating, depression, and increased eating) after implementation of the smoking ban. However, the magnitude of these changes across the two time periods was less than or equal to 0.5 points, and the total withdrawal score did not increase significantly. Control subjects, on the other hand, reported significant decreases over time on five withdrawal items (urges to smoke, irritability, restlessness, impatience, headaches) and a significant (3.23-point) decrease in total withdrawal score (all P values $< .05$). Only scores on the craving-for-cigarettes and urges-to-smoke items were significantly higher in the restricted group vs the control group during weeks 5–8 after the ban ($P < .01$).

Profile of Mood States. The Total Mood Disturbance score, shown in Table 2, had a significant interaction ($F[1, 65] = 5.39, P < .02$) with an increase in the restricted group scores and a decrease in the control group scores. However, these within-group changes were not statistically significant in posthoc testing. Two interactions on the Profile of Mood States factors of Tension-Anxiety ($P < .03$) and Depression-Dejection ($P < .002$) reflected significant preban group differences. In both cases, the control group had higher scores, indicating greater symptom reporting. The restricted group showed no significant changes in factor scores between the preban and the postban periods, although scores on the Fatigue factor decreased significantly over time in the control group ($P < .05$).

Work attitudes and performance. Four of the five measures of work satisfaction and the overall well-being score had significant interactions in the repeated-measures ANOVAs ($P < .03$). Posthoc analyses shown in Table 2 revealed that restricted group members noted small but significant decreases on self-reports of concentration, ($P < .05$), productivity ($P < .01$), relations with coworkers ($P < .05$), and overall well-being ($P < .05$). Control subjects showed no changes over time on the work-productivity indices. Only the overall well-being item had a significant between-group difference after the ban, with the restricted group reporting a lower score ($q = 4.39, P < .05$).

Discussion

This study showed reduced daytime smoking after implementation of a workplace smoking ban. Specifically, smoking during work hours was reduced by an average of four cigarettes per day. This reduction is consistent with the self-report data from Johns Hopkins Hospital smokers[2] and with the results reported for light and moderate smokers in the Australian Public Service.[4] The present findings are also consistent with a recent report of a cross-sectional study by Kinne et al.,[14] which found that men employed in work sites with restrictive smoking policies

smoked fewer cigarettes overall than did men employed in settings without restrictive smoking policies. We found that the decrease in smoking measured at the end of the work shift was corroborated by reductions in the biological measures of salivary nicotine and expired breath carbon monoxide, both of which reflect recent smoking exposure. Further, questions regarding locations where employees smoked revealed the expected shift from smoking at the workstation to smoking outside the building as a result of the smoking ban.

Although it is clear that exposure during the work shift declined, it is important to consider the impact of the ban on overall smoking behavior and tobacco smoke exposure. The most striking finding was that smokers did not compensate for the pack per week reduction either by increased smoking outside of work or by more intensive smoking. Thus, a reduction in overall exposure might be expected. Cotinine levels, which reflect accumulated nicotine exposure over several previous days, declined in the restricted group from 211 to 180 ng/mL, a 15% statistically nonsignificant reduction. Because this reduction suggested an overall trend in reduced exposure, we further examined the data for changes in individual subjects' cotinine levels. Although a full 71% (n = 24) showed some cotinine reduction, only 38% (n = 13) showed a reduction of 50 ng/mL or more. The average postban cotinine level observed among those who had reduced cotinine levels was 167.6 ng/mL (SD = 97.0 ng/mL). This level was only slightly lower than 180 ng/mL, the average postban level for the entire restricted group, and indicated continued moderate levels of smoking exposure despite the trend toward reduction in many subjects.

The increased mean cotinine level observed for control subjects was unexpected (particularly in its magnitude) and was influenced by several individuals with especially large increases (21% of subjects had cotinine increases of > 100 ng/mL). It is unlikely that this observed increase reflects a population tendency toward increasing cotinine levels over time. It is possible that some control subjects reduced their smoking initially as a result of entering a study about smoking behavior.

A larger sample size would have been helpful in assessing the statistical significance of reductions in cotinine levels resulting from a work-site smoking ban. In terms of understanding the clinical significance, however, changes in cotinine levels must be interpreted in light of observed changes in overall tobacco consumption. The work-shift smoking reduction observed was equivalent to one pack of cigarettes per week (4 cigarettes/day × 5 days/week). In a one-pack-per-day smoker, this would amount to a 14% reduction in smoking exposure over a 1-week period. In fact, this smoking reduction is consistent with the observed 15% reduction in cotinine levels and also consistent with the conclusion that there were no compensatory changes in smoking topography or increases in numbers of cigarettes smoked outside of work. Although this is a small reduction in exposure for individual subjects, at the population level, a 15% reduction in tobacco exposure overall could confer an appreciable reduced health risk given that there is strong evidence for a dose-response relationship between smoking and health effects.[15]

Smokers undergoing an abrupt work-site smoking ban showed statistically significant increases in four of the most reliably reported tobacco withdrawal symptoms: difficulty concentrating, craving cigarettes, increased eating, and depression. The subjective report scores of control subjects tended to decrease over time, either because of increases in smoking exposure (as suggested by the cotinine data) or because of a nonspecific tendency for regression to the mean. Restricted smokers reported some difficulties with work productivity and a decline in general well-being. Further, scores on the craving-cigarettes and urges-to-smoke items were significantly higher for the restricted group than for the control group during weeks 5–8 after the ban. Although the magnitude of average change on these measures was quite small, subjects exhibited notable individual differences. The number of restricted smokers who had a mean total withdrawal symptom score of at least "moderate" (> 18) increased by 56% from the preban to the postban period (four subjects at preban and nine at postban). In contrast, for the control group, the number of subjects with moderate withdrawal scores remained steady over time: six and five, respectively. It would be interesting to follow the pattern of the withdrawal symptoms for a longer duration to determine whether there are any lasting effects of work-site smoking restrictions. The importance and increasing popularity of work-site smoking restrictions underscore the need to educate employees and employers alike about the side effects of abstinence during the work shift.

In summary, implementation of a total ban on indoor work-site smoking resulted in a decrease in smoking during the work shift, as verified by reduced

TABLE 2—Preban and Postban Mean Values for Questionnaire Items

Questionnaire Item	Restricted Group (n = 34)		Control Group (n = 33)	
	Weeks 1–4 Mean (SD)	Weeks 5–8 Mean (SD)	Weeks 1–4 Mean (SD)	Weeks 5–8 Mean (SD)
Withdrawal symptom questionnaire				
Difficulty concentrating[c]	0.4 (0.5)	0.7 (0.6)[a]	0.6 (0.7)	0.5 (0.6)
Increased eating	0.3 (0.4)	0.7 (0.9)[a]	0.6 (0.6)	0.7 (0.7)
Depression[c]	0.3 (0.4)	0.5 (0.6)[a]	0.6 (0.6)	0.4 (0.6)
Craving for cigarettes[b,c]	1.3 (0.2)	1.8 (0.3)[a]	1.5 (0.3)	1.2 (0.2)
Urges to smoke[b,c]	1.8 (0.7)	1.9 (0.6)	1.8 (0.5)	1.5 (0.6)[a]
Total withdrawal score[c]	10.1 (6.3)	12.1 (8.2)	13.3 (7.6)	10.1 (7.3)[a]
Total Mood Disturbance score[c]	10.0 (25.9)	14.4 (31.2)	18.8 (31.4)	10.3 (27.1)
Work attitudes and performance				
Ability to concentrate[c]	8.6 (1.3)	7.9 (1.3)[a]	8.2 (1.4)	8.0 (1.5)
Productivity[c]	8.7 (1.4)	8.1 (1.5)[a]	8.1 (1.5)	8.1 (1.5)
Relations with coworkers[c]	9.2 (0.7)	8.5 (1.2)[a]	8.7 (1.1)	8.7 (1.2)
Enjoyment of work[c]	7.6 (1.6)	7.0 (1.7)	7.0 (2.2)	7.2 (2.0)
Overall well being[b,c]	7.9 (1.4)	7.2 (1.7)[a]	7.6 (1.5)	7.9 (1.5)

[a]Significant posthoc test for preban vs postban period.
[b]Significant group differences at postban period.
[c]Significant interaction in analysis of variance.

nicotine and carbon monoxide levels measured at the end of the work shift. The reported reduction in numbers of cigarettes smoked was equivalent to about one pack per week. The cotinine analysis suggested, however, that the smoking ban intervention may provide limited immediate health benefits to smokers themselves in terms of tobacco exposure reduction, while producing minor withdrawal discomfort. Another potential health benefit for smokers that has received some support in the literature is that work-site smoking bans have the additional effect of motivating a small number of smokers to quit completely.[16] However, this effect was not observed during the relatively brief assessment period of this investigation. □

Acknowledgments

This study was supported by two National Institute on Drug Abuse grants, USPHS DA03893 and T32-DA07209.

An earlier version of this article was presented as a Citation Paper at the annual meeting of the Society of Behavioral Medicine, Chicago, Ill, April 18–21, 1990.

We would like to thank an anonymous reviewer who provided insightful commentary and recommendations.

References

1. *The Health Consequences of Involuntary Smoking: A Report of the Surgeon General.* Rockville, Md: Office on Smoking and Health; 1986.
2. Stillman FA, Becker DM, Swank RT, et al. Ending smoking at the Johns Hopkins medical institutions. *JAMA.* 1990;2264: 1565–1569.
3. Becker DM, Conner HF, Waranch R, et al. The impact of a total ban on smoking in the Johns Hopkins Children's Center. *JAMA.* 1989;262:799–802.
4. Borland R, Chapman S, Owen N, Hill D. Effects of work place smoking bans on cigarette consumption. *Am J Public Health.* 1990;80:178–180.
5. Hughes JR, Hatsukami DK. Signs and symptoms of tobacco withdrawal. *Arch Gen Psychiatry.* 1986;43:289–294.
6. Hatsukami DK, Dahlgren L, Zimmerman R, Hughes JR. Symptoms of tobacco withdrawal from total cigarette cessation versus partial cigarette reduction. *Psychopharmacology.* 1988;94:242–247.
7. West RJ, Russell MAH, Jarvis MJ, Feyerabend C. Does switching to an ultra-low nicotine cigarette induce nicotine withdrawal effects? *Psychopharmacology.* 1984; 84:120–123.
8. Russell MAH, Jarvis M, Iyer R, Feyerabend C. Relation of nicotine yield of cigarettes to blood nicotine concentrations in smokers. *Br Med J.* 1980;280:972–976.
9. Henningfield JE, Griffiths RR. Effects of ventilated cigarette holders on cigarette smoking by humans. *Psychopharmacology.* 1980;68:115–119.
10. Benowitz NL, Jacob P, Kozlowski LT, Yu L. Influence of smoking fewer cigarettes on exposure to tar, nicotine, and carbon monoxide. *N Engl J Med.* 1986;315:1310–1313.
11. Young JC, Robinson JC, Rickert WS. A study of chemical deliveries as a function of cigarette butt length. *Beitr Tabakforschung Int.* 1981;11:87–95.
12. Hughes JR, Gust SW, Skoog K, Keenan RM, Fenwick JW. Symptoms of tobacco withdrawal: a replication and extension. *Arch Gen Psychiatry.* 1991;48:52–59.
13. McNair DM, Lorr M, Droppleman LF. *EdITS Manual for the Profile of Mood States.* San Diego, Calif: Educational and Industrial Testing Service; 1971.
14. Kinne S, Kristal AR, White E, Hunt J. Work-site smoking policies: their population impact in Washington State. *Am J Public Health.* 1993;83:1031–1033.
15. *Reducing the Health Consequences of Smoking, 25 Years of Progress: A Report of the Surgeon General.* Rockville, Md: Office on Smoking and Health; 1989.
16. Sorsensen G, Rigotti N, Rosen A, Pinney J, Prible R. Effects of a worksite nonsmoking policy: evidence for increased cessation. *Am J Public Health.* 1991;81:202–204.

Work-Site Cholesterol Screening and Dietary Intervention: The Staff Healthy Heart Project

Alexandra Barratt, MBBS, MPH, Robert Reznik, MD, MBBS, MSc, Les Irwig, PhD, MBBCh, America Cuff, BSc, Judy M. Simpson, PhD, BSc, Brian Oldenburg, PhD, BSc, John Horvath, MBBS, FRACP, and David Sullivan, MBBS, FRACP, FRCPA, for the Steering Committee of the Staff Healthy Heart Project

ABSTRACT

Objectives. The Staff Healthy Heart Project was established to run a work-site cholesterol screening project and a randomized controlled trial of dietary interventions.

Methods. Screening was offered to all staff at six Australian hospitals. Participants with blood cholesterol of 5.2 mmol/L (200 mg/dL) or above were randomly allocated to receive screening only (control group), a self-help package, or a nutrition course. Participants were seen 3 and 6 months after intervention to measure blood cholesterol and dietary changes.

Results. Eighty percent of available staff (n = 2638) were screened. Of those eligible, 67% (n = 683) entered the trial. Follow-up measures of blood cholesterol and dietary intake were obtained for 63% and 38% of trial participants, respectively. A reduction in reported dietary fat was found for all groups, but there were no significant differences between groups. Reported dietary fiber rose by 0.6 g/MJ/day for those in the nutrition course. There were no changes in total blood or high-density lipoprotein cholesterol.

Conclusions. Cholesterol reduction was not demonstrated, but this result is difficult to interpret given the poor ongoing participation rates. Strategies to improve ongoing participation in work-site projects are needed to achieve adequate assessment of dietary interventions used in cholesterol screening. (*Am J Public Health.* 1994;84:779–782)

Introduction

Cholesterol levels are higher than desirable (above 5.2 mmol/L or 200 mg/dL) in approximately half of all adults in Australia and the United States.[1,2] Regular cholesterol checks for all adults have been recommended,[3,4] and work-site cholesterol screening has become a popular strategy for achieving this objective.[5] However, work-site studies often have major methodological problems, including lack of evaluation, and low participation rates.[5,6] Very few work-site cholesterol screening programs have been evaluated by randomized controlled trials, and these have yielded conflicting results, ranging from a net cholesterol reduction of 6.4% in the treatment group in a study of 145 participants[7] to no significant effect in a study of 2489 subjects.[8] Participation rates vary widely (e.g., from 35%[8] to 86%[9]), as do dropout rates (e.g., from 12%[10] to 43%[7]). These problems have made it difficult to draw conclusions about the efficacy of work-site cholesterol screening.

The Staff Healthy Heart Project was established to examine the feasibility of conducting a large work-site cholesterol screening project and to evaluate by randomized controlled trial two dietary interventions to lower cholesterol.

Methods

The project was conducted between 1988 and 1991 at the Royal Prince Alfred Hospital, a tertiary referral center and teaching hospital in Sydney, and at five nearby smaller hospitals. At the outset, a steering committee of 16 members was established to be responsible for the design and implementation of the study. The committee had representatives of senior management staff, the major on-site unions, and the Sports and Social Club at the Royal Prince Alfred Hospital, as well as members representing medical, health promotion, and research interests.

Screening

Two health educators took a mobile screening unit to all major work areas during day and night shifts. Screening was free. Participation was estimated from staff lists supplied by the personnel departments and represents the proportion of staff screened after vacant, "frozen," casual, and off-site positions were deducted from total staff numbers.

Cholesterol was measured by rapid dry chemistry analysis, using the Boehringer-Mannheim Reflotron system. This sample was used for screening purposes only and was available in 3 minutes. Participants' weight and height (without shoes), systolic and diastolic blood pressure (according to standard guidelines[11]), and exercise and smoking habits (elicited by standard questions[1]) were also mea-

Alexandra Barratt, Les Irwig, Judy M. Simpson, and Brian Oldenburg are with the Department of Public Health, University of Sydney, Sydney, NSW, Australia. Robert Reznik is with the Public Health Unit for Central and Southern Sydney. America Cuff, John Horvath, and David Sullivan are with the Royal Prince Alfred Hospital, Sydney.

Requests for reprints should be sent to Alexandra Barratt, MBBS, MPH, Department of Public Health, A27, University of Sydney, Sydney, NSW, Australia 2006.

This paper was accepted September 10, 1993.

Editor's Note. See related editorial by Nestle (p 713) in this issue.

sured. Limited sociodemographic data were collected. Results were reviewed with each participant, and brief advice was given based on pamphlets supplied by the National Heart Foundation. Participants with blood pressure above 140/90 were referred for medical care.

Trial of Dietary Interventions

Eligible employees were those who had cholesterol levels at screening of 5.2 mmol/L or above, were permanent staff, and were fluent in English. Participants were allocated at random in a ratio of 3:3:1 to the control group (screening only) and two intervention groups (self-help package and nutrition course, respectively). For ethical reasons, participants whose screening cholesterol was 7 mmol/L or above were allocated only to intervention groups.

At baseline, venous blood was analyzed for total and high-density lipoprotein cholesterol level at the Biochemistry Department at the Royal Prince Alfred Hospital. All participants were asked to complete a 179-item quantitative food frequency questionnaire (FREQPAN[12]).

Participants were asked (in up to two letters and two phone calls) to attend follow-ups 3 and 6 months after intervention. Blood cholesterol was measured as at baseline, and participants were weighed. A repeat food frequency questionnaire was administered at the 6-month follow-up only. Results of all cholesterol tests were mailed to all participants.

The dietary interventions were consistent with dietary recommendations to reduce total and saturated fat and increase fiber intake.[13] The self-help package, "Food for a Healthy Heart," was a low-intensity intervention. The package included a workbook with educational material and aids to behavior change such as quizzes, shopping guidelines and recipes, a 3-minute video outlining the principles of a healthy heart diet, and a monitoring sheet of suggested dietary behavior changes. The nutrition course, "More Fiber, Less Fat," consisted of five 1-hour sessions led by a dietitian, with 12 to 15 participants per class. The educational content was similar to that of the self-help package, but the sessions provided an opportunity to demonstrate and discuss a standardized series of topics, including the fat and fiber content of foods, food labeling, why people eat too much fat and not enough fiber, and barriers to dietary change. Participants in the sessions were able to taste low-fat, high-fiber recipes. A workbook included information and activities for each session. Participants were consulted about their preferred times to attend and were advised of class times by mail and telephone. Half an hour of paid time off work to attend each class was negotiated.

Participants with a body mass index ([weight in kilograms]/[height in meters]^2) of more than 25 were classified as overweight, and those with a body mass index of more than 30 as obese. The food frequency data were coded and adjusted for serving size according to FREQPAN guidelines.[12] Nutrients were expressed as a proportion of total energy to adjust for differences in energy intake. Changes in nutrient intake, weight, and cholesterol were calculated for each individual at each follow-up attended, and the mean change and 95% confidence interval were calculated for each intervention group. Each intervention group was compared with the control group for significant differences, except for changes in blood cholesterol. For blood cholesterol, comparisons were made to take into account the allocation of people with cholesterol levels of 7 mmol/L or above to intervention groups only (the control group was compared with a subset of the self-help group who had screening cholesterol levels of less than 7 mmol/L [self-help package—low], and the whole of the self-help group was compared with the whole of the nutrition course group). The original allocation to experimental groups was preserved for all analyses. Differences between groups were tested with chi-square and t tests for independent samples for categorical and continuous data, respectively. Two-sided tests of significance at the .05 level were used throughout. The Keys[14,15] and Hegsted[16,17] equations were used to predict the mean change in blood cholesterol from the observed dietary changes for each intervention group.

Results

Results of Screening

A total of 2638 people, approximately 80% of available staff, were screened. The median age was 35 years (range = 15–79, mean = 36.8, SD = 11.5), and 73% of the participants were female. Most (70%) spoke only English; more than 40 other languages were spoken, most commonly Greek, Yugoslav, Chinese languages, Italian, and Spanish.

With respect to cholesterol, 47% of the men and 42% of the women had levels of 5.2 mmol/L or above, and 7% of the men and 5% of the women had levels of 7 mmol/L or above. Almost half (47%) of the men, and 36% of the women, were overweight; 10% of the men and 12% of the women were obese. Over half the sample (53% of the men and 57% of the women) reported engaging in no regular vigorous exercise (e.g., football, keep-fit classes) in the preceding 2 weeks. A quarter of the women and 30% of the men reported current smoking. Only a small proportion (7% of the men and 3% of the women) were found to have high blood pressure (diastolic pressure above 90 mm Hg).

Results from the Trial of Dietary Interventions

Of those eligible for the trial, 67% (n = 683) agreed to participate. Refusers were significantly more likely to be male and to be smokers and were on average 2 years younger. The participants were allocated into groups as follows: screening only (control), 259; self-help, 310; and nutrition course, 114. A baseline food frequency questionnaire was obtained from 79% of the participants. Of participants allocated to the nutrition course, 38 (33%) failed to attend any classes, 40 (35%) attended three or more classes, and 36 (31%) attended one or two classes.

Follow-up cholesterol samples were obtained from 61% and 63% of the participants at the 3- and 6-month follow-ups, respectively. Repeat measures of weight were obtained from 51% (at 3 months) and 49% (at 6 months) of the participants. Compared with those seen at follow-up, dropouts were significantly heavier and had higher blood pressure at screening, but the two groups were similar with respect to gender, age, mean baseline cholesterol level, exercise, and smoking. A repeat food frequency questionnaire was completed by 48% of the trial participants; paired baseline and follow-up food frequency questionnaires were available for 261 (38%) of the participants. Participants who failed to complete both food frequency questionnaires were younger (3 years on average) and were more likely to be male and to smoke.

Of the 261 paired questionnaires, 10 were excluded from analysis because of extreme values of reported energy intake (outside the 1st and 99th percentiles) or extremely large drops (greater than 10 MJ) in energy intake.

Compared with the other groups, the nutrition course significantly reduced reported total energy intake and increased fiber intake (Table 1). A reduction in total and saturated fat was reported in all

TABLE 1—Baseline Intakes of Energy, Total Fat, Saturated Fat, and Fiber and Change at 6-Month Follow-Up, by Intervention Group

Group	Energy, MJ		Total Fat, % of Total Energy		Saturated Fat, % of Total Energy		Fiber, g/MJ	
	Baseline	Change at 6 Months	Baseline	Change at 6 Months	Baseline	Change at 6 Months	Baseline	Change at 6 Months
Screening only	9.2	−1.1[a]	36.8	−2.0[a]	13.8	−1.0[a]	2.9	+0.1
Self-help package	8.7	−0.9[a]	37.3	−2.4[a]	14.2	−1.7[a]	2.7	+0.2[a]
Nutrition course	10.0	−2.1[a,b]	37.5	−2.5	13.9	−1.7	2.7	+0.6[a,c]

Note. Paired baseline and follow-up questionnaires were obtained for 261 participants (38% of the total n of 683). Ten questionnaires were excluded from analysis because of extreme values; therefore, the table presents data for 251 participants.
[a]95% confidence interval does not include zero.
[b]Change in nutrition course group is significantly different from change in self-help package group ($P = .04$) and change in screening-only group ($P = .05$).
[c]Change in nutrition course group is significantly different from change in screening-only group ($P = .04$).

groups, but the reduction was not significantly greater in either intervention group than in the control group. There were no changes in intake of polyunsaturated or monounsaturated fats.

There were no changes in mean cholesterol level in any group at the 3- or 6-month follow-up (Table 2). The maximum possible decrease in cholesterol (from the lower bound of the 95% confidence interval) was no more than approximately 4% for any group. There were no changes in high-density lipoprotein cholesterol. The maximum decrease in cholesterol predicted from the observed dietary changes was 0.12 mmol/L (2%) for those in the nutrition course.

A mean weight loss of 0.9 kg for the nutrition course group was observed at the 3-month follow-up, but this change was not significantly different from weight changes in the other groups. At 6 months participants in the nutrition course continued to show a 0.25-kg mean weight loss, a statistically significant loss compared with that of the screening-only group ($P = .04$).

TABLE 2—Baseline Cholesterol Levels (mmol/L) and Changes at 3 and 6 Months, by Intervention Group

Group	Baseline Level (n = 668)	Change at 3 Months (n = 417)		Change at 6 Months (n = 430)	
		Mean	95% CI	Mean	95% CI
Screening only	5.82	−0.07	−0.19, +0.05	−0.10	−0.22, +0.01
Self-help package—low	5.79	−0.05	−0.16, +0.05	−0.01	−0.10, +0.09
Self-help package	5.99	−0.05	−0.15, +0.04	−0.02	−0.10, +0.07
Nutrition course	6.16	+0.06	−0.14, +0.26	−0.05	−0.25, +0.15

Note. CI = confidence interval. The control group was compared with the subset of participants in the self-help package group whose screening cholesterol was below 7.0 mmol/L (self-help package—low); the whole of the self-help package group was compared with the whole of the nutrition course group (see Methods).

Discussion

An effective union between academic medicine and the work site is needed to exploit the full potential of work-site health promotion.[18] In the Staff Healthy Heart Project a major effort was made to achieve an effective collaboration among members of the steering committee. Broad agreement on the risk factors to be addressed and the interventions to be used was achieved among researchers, management, and unions. Management and union representatives were then able to negotiate agreements on provision of time off work and use and confidentiality of data. The head of each department was asked personally to support the project, to allocate a room for screening, and to allow staff time away from their duties to participate. This cooperation among researchers, management, and staff was probably responsible for the high participation in screening. The participation rate of about 80% compares favorably with those of other work-site trials, including the World Health Organization European trial in which participation rates ranged from 75% to 89%.[9]

The recruitment rate (67% of eligible staff) compares well with rates of 50% or less in other work-site trials of cholesterol intervention.[7,8] Women, older participants, and nonsmokers were more likely to be recruited. However, the relatively high recruitment rate in this population may have resulted in the inclusion of participants who were not sufficiently motivated to complete the trial. In contrast, Edye et al. recruited only 35% to 45% of participants initially, but retained almost 80% after 3 years.[8] These authors also noted that young women were more likely to drop out. Our study population was unusual in that it was predominantly young and female, and this characteristic may have increased the difficulty of achieving ongoing participation.

Organizational factors are also considered important determinants of participation.[19] Although screening was very convenient, follow-up was less convenient. Participants had to leave their work areas and walk to the staff clinic; follow-up took about 15 minutes at 3 months and about 45 minutes at 6 months (to complete the food frequency questionnaire in addition to the other measures). Participants also had to leave their work areas to attend the nutrition sessions, and many reported difficulty in leaving because of the pressure of work, even though time off work to attend had been arranged. However, in another work-site trial of nutritional classes, the proportion attending no classes (32%) was similar to that in our trial.[7] Follow-up rates were also reduced by the high staff turnover at

this work site; approximately one quarter of those who did not participate in follow-up had resigned or retired.

Finally, follow-up rates varied by outcome measure and were worst for the food frequency questionnaire. The baseline response rate was 79% (similar to that reported by other authors using food frequency questionnaires[20,21]), so the follow-up rate was limited to a proportion of these participants. The questionnaire was quite long and detailed and some participants were reluctant to spend the time needed to complete it, particularly for a second time. However, as no brief, valid, self-report measure of fat intake was available, we chose the food frequency questionnaire as the most appropriate and practical form of dietary assessment for this study.

Each intervention group reported appropriate dietary changes, including reductions in total and saturated fat intake. In addition, participants in the nutrition course increased their fiber intake and reduced their weight, compared with the control group. No benefit for intervention groups in terms of cholesterol reduction was demonstrated, although this result should be interpreted with caution in view of the low ongoing participation rates.

Conclusions

This project demonstrated that a large work-site screening program is feasible and high rates of initial participation can be achieved. The dietary interventions achieved some self-reported dietary change, but no reduction in blood cholesterol was demonstrated. However, this result is difficult to interpret because of the low rates of ongoing participation in the trial. Strategies to maintain high rates of participation beyond the initial stages of work-site projects are required. These strategies might include ways of making programs more accessible and convenient for staff. Such strategies would make a major contribution to the evaluation of dietary interventions used in cholesterol screening programs. □

Acknowledgments

Funding for this project was received from the Central Sydney Area Health Service, Royal Prince Alfred Hospital, the National Heart Foundation of Australia, and the National Health and Medical Research Council (through the Public Health Research and Development Committee). Additional sponsorship was received from the Australian Nutrition Foundation; Pfizer Pty Ltd; Becel Heart Care Information Service; Merck, Sharp & Dohme; SP Agencies; and Reebok Australia.

We would like to acknowledge the contribution made to this project by Dr Sue Morey, Chief Medical Officer, New South Wales Department of Health (formerly Director, Department of Community Medicine, Royal Prince Alfred Hospital) in establishing the project, and by the other members of the Steering Committee—Prof Douglas Baird, Chairman, Central Sydney Area Health Service; Ms Jo Rogers, Australian Nutrition Foundation; Dr Phil Harris; Dr Ian Caterson; Ms Kate Hogan; Ms Marilyn Wise; Ms Vanessa Child; Ms Katherine Maxwell; Mr Terry Conlon; and Ms Annabelle Harman of Royal Prince Alfred Hospital—who have provided much support and assistance.

References

1. Risk Factor Prevalence Study Management Committee. *Risk Factor Prevalence Study: Survey No. 3, 1989.* Canberra, ACT, Australia: National Heart Foundation of Australia and Australian Institute of Health; 1990.
2. Sempos CT, Cleeman JI, Carroll MD, et al. Prevalence of high blood cholesterol among US adults: an update based on guidelines from the second report of the National Cholesterol Education Program Adult Treatment Panel. *JAMA.* 1993;269:3009–3014.
3. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Summary of the Second Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II). *JAMA.* 1993;269:3015–3023.
4. The management of hyperlipidaemia: a consensus statement. *Med J Aust.* 1992; 156(suppl):S1–S8.
5. Fielding JE. Frequency of health risk assessment activities at U.S. worksites. *Am J Prev Med.* 1989;5(2):73–81.
6. Sallis JF, Hill RD, Fortmann SP, Flora JA. Health behavior change at the worksite: cardiovascular risk reduction. In: Hersen M, Eisler EM, Miller PM, eds. *Progress in Behavior Modification.* New York, NY: Academic Press; 1986;20:161–197.
7. Bruno R, Arnold C, Jacobson L, Winick M, Wynder E. Randomized controlled trial of a nonpharmacologic cholesterol reduction program at the worksite. *Prev Med.* 1983;12: 523–532.
8. Edye BV, Mandryk JA, Frommer MS, Healey S, Ferguson D. Evaluation of a worksite programme for the modification of cardiovascular risk factors. *Med J Aust.* 1989;150:574–581.
9. World Health Organization European Collaborative Group. Multifactorial trial in the prevention of coronary heart disease, 1: recruitment and initial findings. *Eur Heart J.* 1980;1:73–80.
10. Lovibond SH, Birrell PC, Langeluddecke P. Changing coronary heart disease risk factor status: the effects of three behavioral programs. *J Behav Med.* 1986;9:415–436.
11. *Guidelines on Detection of High Blood Pressure and its Evaluation.* Canberra, ACT, Australia: National Heart Foundation of Australia; 1987.
12. Record S, Baghurst K. *Dietary Assessment by Computer.* Adelaide, SA, Australia: Commonwealth Scientific and Industrial Research Organisation; 1987.
13. The Health Targets and Implementation (Health for All) Committee. *Health for All Australians.* Canberra, ACT, Australia: Australian Government Publishing Service; 1988.
14. Keys A, Parlin RW. Serum cholesterol response to changes in dietary lipids. *Am J Clin Nutr.* 1966;19:175–181.
15. Keys A. Serum cholesterol response to dietary cholesterol. *Am J Clin Nutr.* 1984;40: 351–359.
16. Hegsted DM, McGandy RB, Myers ML, Stare FJ. Quantitative effects of dietary fat on serum cholesterol in man. *Am J Clin Nutr.* 1965;17:281–295.
17. Hegsted DM. Serum cholesterol response to dietary cholesterol: a re-evaluation. *Am J Clin Nutr.* 1986;44:299–305.
18. Alderman MH. Academic medicine and the workplace. *Am J Public Health.* 1993;83: 313–314.
19. Sloan RP, Gruman JC. Participation in workplace health promotion programs: the contribution of health and organizational factors. *Health Educ Q.* 1988;15:269–288.
20. Baghurst KI, Crawford DA, Worsley A, Record SJ. The Victorian Nutrition Survey—intakes and sources of dietary fats and cholesterol in the Victorian population. *Med J Aust.* 1988;149:12–20.
21. Hebert JR, Harris DR, Sorensen G, Stoddard AM, Hunt MK, Morris DH. A work-site nutrition intervention: its effects on the consumption of cancer-related nutrients. *Am J Public Health.* 1993;83:391–394.

Improving Dietary Behavior: The Effectiveness of Tailored Messages in Primary Care Settings

Marci Kramish Campbell, PhD, RD, MPH, Brenda M. DeVellis, PhD, Victor J. Strecher, PhD, MPH, Alice S. Ammerman, DrPH, RD, Robert F. DeVellis, PhD, and Robert S. Sandler, MD, MPH

ABSTRACT

Objectives. To achieve the *Healthy People 2000* objectives, public health professionals must develop effective dietary interventions that address psychosocial and behavioral components of change. This study tested the effect of individually computer-tailored messages designed to decrease fat intake and increase fruit and vegetable intake.

Methods. Adult patients from four North Carolina family practices were surveyed at baseline and then randomly assigned to one of two interventions or to a control group. The first intervention consisted of individually computer-tailored nutrition messages; the second consisted of nontailored nutrition information based on the 1990 *Dietary Guidelines for Americans*. Patients were resurveyed 4 months postintervention.

Results. The tailored intervention produced significant decreases in total fat and saturated fat scores compared with those of the control group ($P < .05$). Total fat was decreased in the tailored group by 23%, in the nontailored group by 9%, and in the control group by 3%. Fruit and vegetable consumption did not increase in any study group. Seventy-three percent of the tailored intervention group recalled receiving a message, compared with 33% of the nontailored intervention group.

Conclusions. Tailored nutrition messages are effective in promoting dietary fat reduction for disease prevention. (*Am J Public Health.* 1994;84:783–787)

Introduction

The *Healthy People 2000 Objectives for the Nation* recommend that Americans decrease dietary fat to 30% or less of their total caloric intake and increase fruit and vegetable consumption to five or more daily servings in order to promote health and prevent diet-related diseases, including heart disease and certain cancers.[1] To achieve these objectives, public health professionals must find effective methods of integrating behavioral and psychosocial factors into nutrition-related behavior change programs. Because most behavior change interventions are designed for individuals who are already prepared for action, they may be inappropriate for those not yet ready to act.[2]

The Partners in Prevention–Nutrition program was designed to influence dietary intake by providing mailed, individually computer-tailored messages that address dietary and psychosocial factors and provide specific strategies for change. The intervention used a stage-of-change approach, based on the Prochaska and DiClemente transtheoretical framework,[3] to match tailored communications to the needs of participants at different points in the decision process for dietary change. We hypothesized that providing information relevant to an individual's stage of change would promote movement from earlier to more advanced stages of readiness, ultimately leading to actual behavior change. Accordingly, we conducted a randomized trial using a pretest and posttest to assess the impact of tailored vs nontailored nutrition education materials on fat, fruit, and vegetable consumption.

Methods

Study Population

Respondents were 558 adult patients (ages 18 and above) recruited from four family practices in central North Carolina between September and November 1991. Two practices served a primarily urban population and two were primarily rural. Family practice office staff recruited participants as they checked in for any type of medical appointment. Office staff excluded patients who were too ill or mentally unable to complete the baseline survey. Only one family member per household was allowed to participate.

The study sample was selected from an initial pool of 682 individuals who met eligibility criteria. A total of 124 people (18.2%) were excluded during the baseline survey phase for the following reasons: refusals (8.8%), consent form not signed (3.7%), ineligible due to age (< 18) or multiple household membership (2.2%), survey incomplete (3.4%), and participation in a related study (one individual). An additional 95 participants (13.9%) were lost during the follow-up survey phase.

Procedures

Participants were assessed initially regarding fat, fruit, and vegetable dietary intake; gender; age; race; education; marital status; physician-prescribed diets; pregnancy status (women only); stage of readi-

Marci Kramish Campbell, Brenda M. DeVellis, Victor J. Strecher, and Robert F. DeVellis are with the Department of Health Behavior and Health Education, School of Public Health; Alice S. Ammerman is with the Department of Nutrition, School of Public Health; and Robert S. Sandler is with the Center for Gastrointestinal Biology and Disease, School of Medicine, at the University of North Carolina, Chapel Hill.

Requests for reprints should be sent to Marci Kramish Campbell, PhD, RD, MPH, Department of Nutrition, School of Public Health, Campus Box #7400, University of North Carolina, Chapel Hill, NC 27599-7400.

This paper was accepted January 10, 1994.

Editor's Note. See related editorial by Nestle (p 713) in this issue.

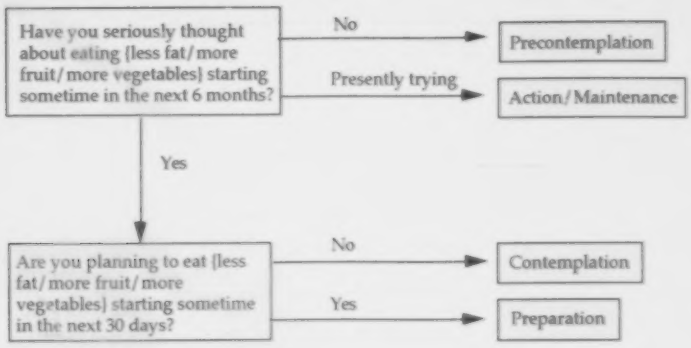


FIGURE 1—Stages of dietary behavior change.

ness to change diet; and psychosocial factors associated with dietary behavior change, such as self-efficacy (confidence in one's ability to change). The self-administered surveys were distributed and collected by family practice staff. Nearly all patients (99%) were able to complete the survey while in the family practice office; those who could not were allowed to take it home and return it using a postage-paid envelope. Participants were then randomly assigned to one of three groups: an intervention group, which received tailored nutrition messages; a comparison intervention group, which received nontailored nutrition messages; and a control group, which did not receive nutrition messages. All messages were mailed to participants within 3 weeks of baseline survey completion.

Participants were resurveyed 4 months postintervention to assess change. Household income also was assessed at follow-up. Through the use of a combined mail and telephone strategy, a final response rate of 82.3% (n = 459) of baseline participants was achieved. Telephone interviewers were blinded to participants' study group membership. Reasons for nonresponse included refusals (3.2%), passive refusals[4] (7.9%), no answer after 15 attempts (0.2%), and disconnected phone or no telephone number provided (5.7%). Four individuals (0.7%) were not resurveyed because one had moved with no forwarding address and three were found to be members of households that included another study participant. Respondents to the follow-up survey were more likely to be married than were nonrespondents; no other demographic or dietary differences distinguished respondents from nonrespondents.

Measures

Food frequency. A food frequency instrument consisting of 28 items derived from the Health Habits and History Questionnaire[5] was developed for this study. Items assessing fat intake included 13 items validated as a brief fat screener[6] plus 5 items representing high-fat foods commonly consumed by southern US populations. Ten items were used to assess fruit and vegetable consumption. Response options were presented as a checklist, with seven response choices: three or more times per day, two per day, about every day, two to four per week, once a week, one to three times per month, and rarely or never. Dietary fat and saturated fat scores were obtained by multiplying frequency of consumption (calculated as servings per day) by portion data for each item[5,7] and summing the items. Fruit and vegetable combined intake was obtained by calculating the frequency of consumption of each item and summing the frequencies to determine total daily servings.

Stage of change. Stage of readiness to change diet was measured for three consumption behaviors: decreasing fat, increasing vegetables, and increasing fruit. Based on a series of questions derived from the Prochaska and DiClemente stages of change for smoking cessation[8] (Figure 1), participants were categorized at baseline as being in one of the following stages: precontemplation (not seriously thinking about change), contemplation (seriously thinking about change within the next 6 months), preparation (planning to change within the next 30 days), or action/maintenance (currently trying to change). The stages-of-change model defines maintenance as continuing behavior change for longer than 6 months, whereas action includes behavior change efforts of less than 6 months.[8] Action and maintenance stages were combined here because our study design did not allow for long enough follow-up to assess maintenance of change. Because 91% of contemplators were planning to change within the next 30 days, the contemplation and preparation stages were also combined into one category (contemplation) for data analysis.

Psychosocial issues. Motives for change and barriers to change were assessed using questions derived from pretesting. Self-efficacy regarding ability to increase fruits and vegetables and to decrease dietary fat was measured using a 4-point scale (1 = very confident, 4 = not at all confident). For fat reduction, three subbehaviors were measured (lowering meat fat, dairy fat, and fats added to foods). Beliefs about perceived susceptibility to diet-related diseases and perceived benefit of dietary change to avoid health problems were also measured using a 4-point scale (1 = very likely, 4 = not at all likely). Participants contemplating change were questioned about past efforts to change and their continuation or relapse after those efforts. These psychosocial variables were measured for use in providing individualized computer-tailored feedback.

Intervention

The tailored intervention consisted of a one-time, mailed nutrition information packet tailored to the participant's stage of change, dietary intake, and psychosocial information. The health behavior literature indicates that tailoring can be an effective tool to address specific psychosocial issues and barriers to change.[9] Through the use of a tailored approach, individuals can be given behavior change information relevant to their stage in the decision process, whereas information not likely to be helpful at that stage can be omitted.

The tailored messages were created with Authorware Professional and Microsoft Word software for the Apple Macintosh computer. A library of text messages was developed to target each survey response choice regarding stage of change, motives, barriers, beliefs, and self-efficacy for each dietary behavior.[10] A team of nutritionists and health educators created, pretested, and revised the messages. Computer algorithms were developed to

access the appropriate text pieces based on each individual's baseline information. The computer then merged the text pieces into tailored nutrition education newsletters using a predetermined overall format.

Baseline food frequency information was used to provide individualized feedback about whether dietary change was needed. For dietary fat, participants were given feedback in graphic form depicting their current intake as "high" or "low-moderate." This evaluation was made by comparing the individual's dietary score with the 75th percentile of the distribution of fat intake in this population, as determined by pretesting with 55 family practice patients prior to the study. For fruits and vegetables, baseline information was used to show the number of daily servings the individual consumed compared with the recommended five or more servings per day.

Messages were structured according to a framework based on the stages of change and the Health Belief Model.[11] All tailored group members were mailed a packet containing a nutrition profile summarizing their current diet and level of interest in changing behavior, a tailored page regarding dietary fat intake, and a tailored page regarding fruit and vegetable intake. Each message acknowledged the participant's stage and addressed his or her beliefs about both susceptibility to diet-related diseases and perceived benefits of and motives for changing diet. Individualized diet feedback was then provided regarding baseline fat and fruit/vegetable intake. Contemplators received information designed to decrease barriers to change and to increase self-efficacy. Depending on stage of change, self-efficacy, and history of past relapse, participants also received tailored recipes and specific diet tips designed to promote skills and to provide cues to action. Those individuals who were already trying to change received tailored recipes and messages aimed at preventing relapse.

The nontailored messages appeared nearly identical to the tailored messages in format, typeface, and paper; however, no individualized dietary or psychosocial information was given. Instead, these messages provided standard risk information about the relation of diet to disease and gave dietary recommendations based on the 1990 *Dietary Guidelines for Americans*.[12] Control group members completed the baseline and follow-up surveys but did not receive any nutrition information from the Partners in Prevention–Nutrition program.

TABLE 1—Family Practice Patients' Baseline Dietary Intake, by Demographic Group and Stage of Change

		Fat (g/day)		Saturated Fat (g/day)		Vegetable/Fruit (servings/day)	
	No.	Mean (SE)	*P**	Mean (SE)	*P**	Mean (SE)	*P**
Gender							
Male	138	51.9 (3.1)	<.001	21.5 (1.3)	<.001	2.9 (.20)	.142
Female	420	43.4 (2.2)		17.3 (1.0)		3.3 (.20)	
Race							
White	448	47.0 (2.2)	.771	19.2 (1.0)	.974	3.0 (.15)	.588
Non-White	106	47.9 (3.3)		19.3 (1.4)		3.2 (.25)	
Education, y							
12 or less	204	50.9 (2.9)	<.01	20.7 (1.2)	<.01	2.8 (.19)	<.001
>12	348	44.2 (2.4)		17.8 (1.1)		3.5 (.21)	
Age, y							
<31	137	56.3 (3.4)	<.001	20.3 (1.2)	<.001	2.9 (.20)	<.005
31–38	143	47.5 (2.8)		19.5 (1.4)		2.9 (.23)	
39–48	141	43.4 (2.6)		17.5 (1.2)		2.8 (.22)	
49+	131	44.3 (3.0)		16.9 (1.2)		3.8 (.30)	
Marital status							
Married	348	49.4 (2.5)	.074	20.1 (1.2)	.072	3.3 (.20)	.117
Nonmarried	204	45.6 (2.6)		18.4 (1.1)		2.9 (.20)	
Stage[a]							
Precontemplation	141	47.5 (2.4)	<.05	19.3 (.96)	<.01	3.1 (.19) (n = 156)	<.005
Contemplation	253	44.7 (1.8)		17.8 (.71)		3.1 (.15) (n = 217)	
Action	164	40.0 (2.0)		15.7 (.79)		3.9 (.24) (n = 185)	

[a]As shown, numbers of participants in each stage of change differed for reducing fat versus increasing fruits and vegetables. The table shows vegetable/fruit-combined consumption according to stage of change for increasing vegetables. The same variables were significant when vegetables and fruit were analyzed separately by respective stage of change.
**P* values for dietary differences are based on *F* tests after adjustment for all effects in the model.

Analysis

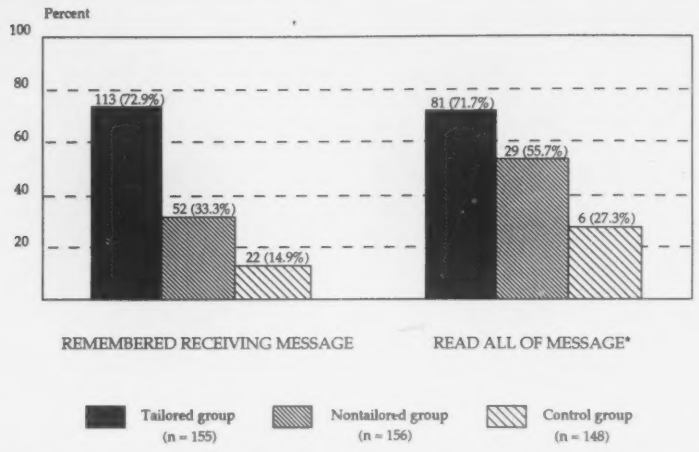
All dietary variables showed a skewed distribution and were transformed to conform more closely with a normal distribution. For fat data, least squares adjusted means were calculated after transformation by the natural logarithm ($e^{\bar{z}}$, where $\bar{z}$ = ln fat or ln saturated fat); standard error was approximated by SE = $(e^{\bar{z}})SE(\bar{z})$. For fruit and vegetable data, least squares adjusted means were calculated from square root transformed variables ($\bar{z}^2$, where $\bar{z}$, = square root daily servings); standard error was approximated by SE = $(\bar{z})^2SE(\bar{z})$. Analysis of covariance was used to test between-group differences for continuous variables, and logistic regression analysis was used to test differences among categorical variables.

Results

Descriptive Results

The study sample was primarily female (75.3%), married (62.3%), and well educated (mean = 13.6 years), with an average age of 40.8 years (Table 1). Minority enrollment was 19.0% of the sample, nearly all African Americans. Of 401 participants who reported their income (87.4% of the follow-up sample), the median annual household level was $30 000 to $39 000. There were no significant differences in demographic, dietary, or stage-of-change variables among study groups. The tailored intervention group consumed more dietary fat at baseline compared with the other two groups; this difference was not statistically significant.

We examined the relationship between stage of dietary change and baseline dietary intake scores, controlling for demographic variables (see Table 1). Self-reported stage of change was a significant predictor of baseline intake. Precontemplators and younger participants consumed significantly more total fat and saturated fat, and fewer fruits and vegetables, than did individuals in the action stage and older participants. Similarly, men and participants with up to 12 years of education consumed more fat and



Note. There are differences in remembering having received the message between the three study groups, $P < .0001$. There are also differences in message readership between the tailored and nontailored study groups, $P < .05$.

FIGURE 2—Differences between study groups in message recall and amount read.

TABLE 2—Comparison of Tailored Message, Nontailored Message, and Control Groups' Reported Consumption of Fat, Saturated Fat, and Fruits/Vegetables at 4 Months Postintervention

	Fat (g/day)		Saturated Fat (g/day)		Vegetable/Fruit (servings/day)	
	Mean (SE)	P[a]	Mean (SE)	P[a]	Mean (SE)	P[a]
Tailored group (n = 134)						
Baseline	45.6 (2.6)	.033	18.7 (1.1)	.036	3.6 (.19)	.817
Follow-up	35.3 (1.7)		13.9 (.72)		3.3 (.19)	
Difference[b]	−10.3		−4.8		−0.3	
Nontailored group (n = 136)						
Baseline	40.4 (2.4)	.157	16.1 (.93)	.110	3.6 (.20)	.968
Follow-up	36.8 (1.7)		14.4 (.72)		3.3 (.19)	
Difference[b]	−3.6		−1.7		−0.3	
Control group (n = 124)						
Baseline	41.1 (2.1)	...	16.3 (.98)	...	3.6 (.20)	...
Follow-up	39.8 (1.9)		15.8 (.81)		3.3 (.20)	
Difference[b]	−1.3		−0.5		−0.3	

Note. Group means are adjusted for baseline intake and the demographic variables of gender, age, education, race, marital status, and income. After deletion of individuals with missing demographic information, n = 394 in the final sample.

[a]*P* values for dietary differences are based on *t* tests after analyses of covariance, comparing each intervention group with the control group.

[b]Difference is computed by subtracting follow-up adjusted mean from baseline adjusted mean. Minus signs are shown to indicate the direction of change.

fewer fruits and vegetables than did women and participants with higher education.

Intervention Effects

As shown in Figure 2, study group (tailored message, nontailored message, or control) was a significant predictor of remembering having received a message ($\chi^2 = 116.35, P < .0001$). A total of 72.9% of individuals in the tailored group recalled receiving nutrition information, compared with 33.3% of the nontailored group and 14.9% of the control group (to whom no message was sent). Study group also predicted message readership among those who remembered having received a message. Members of the tailored group were more likely to report having read all the message than were those in the nontailored group ($\chi^2 = 3.98, P < .05$).

We used analysis of covariance to test the effect of study group on decrease in total fat and saturated fat intake scores while controlling for baseline intake and demographic variables (see Table 2). An overall test of study group effect was marginally significant ($F = 2.37, P < .10$), indicating that differences existed among groups. We then performed contrasts using *t* tests to examine between-group differences. For both total fat and saturated fat, the difference was statistically significant for the tailored group–control group comparison ($t = -2.14, P < .05$ for total fat; $t = -2.11, P < .05$ for saturated fat) but not for the nontailored group–control group comparison. Total fat decreased by 23% (10.3 g) in the tailored group, compared with a decrease of 9% (3.6 g) in the nontailored group and 3% (1.3 g) in the control group. The tailored group decreased saturated fat by 26%, compared with decreases of 11% in the nontailored group and 3% in the control group.

Analysis of covariance was also used to examine study group effect on increase in fruit and vegetable consumption while controlling for baseline intake and demographic variables (Table 2). Study group did not predict an increase in fruit and vegetable intake; consumption decreased by approximately one quarter of a daily serving in all groups.

Discussion

This randomized trial demonstrated a positive effect of tailored nutrition messages in promoting dietary change efforts for disease prevention. Participants who were sent tailored messages were more than twice as likely as those who were sent nontailored messages to remember receiving the information. Those who remembered receiving a message were more likely to report having read all of it if the message was tailored than if it was not tailored. At posttest, participants in the tailored intervention group had reduced their total fat and saturated fat intakes significantly compared with participants in the control group. Members of the nontailored group had reduced intake also; however, the level of change was not significant when compared with that of the control group.

Use of dietary change as an outcome measure assumes that diet modification

will reduce disease risk in the future. The *Healthy People 2000* objectives[1] recommend lowering dietary fat intake to 30% or less of total calories, compared with the current 36% to 38% of calories consumed by most Americans. This corresponds to a recommended reduction in fat intake of about 17% to 21%. In the present study, the tailored message group reduced their fat score by approximately 23% from baseline intake, suggesting that they achieved a level of change commensurate with the *Healthy People 2000* objectives. It is possible, however, that by using a brief assessment instrument, significant sources of dietary fat were missed, and this could have biased the results. The food frequency instrument was designed for self-administration by patients waiting for their physician appointment, and it included only 18 items to measure dietary fat. This instrument did not collect enough information to determine the percentage of calories consumed as fat, which would have provided a more valid and reliable estimate of dietary change. Attempts were made to reduce possible bias by using a randomized design and controlling for baseline intake in the analysis.

The descriptive findings show that the stage-of-change framework can be applied to dietary behaviors in that self-reported stage information was associated with baseline dietary intake. Precontemplators consumed the most fat and the fewest fruits and vegetables at baseline, whereas individuals in the action stage consumed the least dietary fat and the most fruits and vegetables. These results are consistent with other research demonstrating the relation of stage of change to percentage of calories consumed as fat.[13] Our findings suggest that the stages of change are also related to fruit and vegetable dietary behavior.

The observed lack of an intervention effect on fruit and vegetable consumption may have been partly due to seasonal factors. Baseline surveys were completed in the fall, and follow-up surveys were completed in the early spring with respondents being asked to recall the preceding 3 (winter) months. Seasonal variation in fruit and vegetable intake has been documented in other studies.[14] A number of participants commented during the telephone interviews that they planned to increase their intake of fruits and vegetables but were waiting until those foods came into season.

A potential bias in this study is the possibility of higher demand characteristics of the tailored messages, compared with the nontailored information. The tailored messages were personalized and provided more information than did the nontailored messages; therefore, tailored group members may have felt a stronger obligation to report positive changes at follow-up. We attempted to minimize the impact of demand characteristics, however, by designing both intervention materials similarly in terms of layout, logo, type, and paper; by sending personalized cover letters from the family physicians with all follow-up surveys; and by blinding telephone interviewers to study group membership.

The study findings have implications for public health and primary care practice. Individualized health and nutrition guidance has traditionally been provided by physicians, nurses, or nutritionists in the form of individual patient counseling, which is costly, time-consuming, and usually offered only to patients at high risk of diet-related morbidity. Standard practice is to provide nontailored written materials to patients. The present study shows that family practice patients are willing to participate in nutrition education programs designed to prevent chronic diseases, and that computer technology can be used to provide individually tailored information that is more effective than standard educational materials in promoting behavior change. Because computer-tailored education is relatively inexpensive compared with individual counseling, it can be provided to low-risk as well as to high-risk individuals who are interested in improving their health.

This study was conducted in a family practice population. Individuals who obtain care from family physicians may differ from the general public in sociodemographic characteristics and in motivation to participate in health promotion activities. The study sample was of higher education and income status than average for North Carolina. Prior research suggests that tailored mammography screening messages are more effective than nontailored messages for lower-income and minority women.[15] Future research should evaluate the impact of tailored messages in low-literacy populations. □

Acknowledgments

This research was supported by a Cancer Prevention Research Grant from the Lineberger Comprehensive Cancer Center and a grant from the Institute of Nutrition of the University of North Carolina.

The results of this study were presented, in part, at the annual meeting of the American Public Health Association in Washington, DC, on November 11, 1992.

The authors thank David Farrell, MPH, and the University of North Carolina Health Communications Research Laboratory for their contributions to this research. The authors also thank the physicians, staff, and patients of Lakewood on Broad Family Practice, Triangle Family Practice, East Chatham Medical Center, and Orange Family Medical Group.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50212.
2. Prochaska J. A transtheoretical model of behavior change: implications for diet interventions. In: Henderson M, Bowen D, DeRoos K, eds. *Promoting Dietary Change in Communities: Applying Existing Models of Dietary Change to Population-Based Interventions.* Seattle, Wash: Fred Hutchinson Cancer Research Center; 1992:37–50.
3. Prochaska J, DiClemente C, Norcross J. In search of how people change: applications to addictive behaviors. *Am Psychol.* 1992;47: 1102–1114.
4. Mishra SI, Dooley D, Catalano R, Serxner S. Telephone health surveys: potential bias from noncompletion. *Am J Public Health.* 1993;83:94–99.
5. Block G, Hartman A, Dresser C, et al. A data-based approach to diet questionnaire design and testing. *Am J Epidemiol.* 1986; 124:453–469.
6. Block G, Clifford C, Naughton M, Henderson M, McAdams M. A brief dietary screen for high fat intake. *J Nutr Educ.* 1989;21: 199–207.
7. Pennington J, Church H. *Bowes and Church's Food Values of Portions Commonly Used.* 15th ed. Philadelphia, Pa: JB Lippincott; 1989.
8. Prochaska J, DiClemente C. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol.* 1983;51:390–395.
9. Becker M, Maiman L. Strategies for enhancing patient compliance. *J Community Health.* 1980;6:113–135.
10. Campbell MK. *The Impact of Message Tailoring on Nutrition Behavior Change for Disease Prevention.* Chapel Hill, NC: University of North Carolina; 1992. Dissertation.
11. Becker M, ed. The health belief model and personal health behavior. *Health Educ Monogr.* 1974;2:324–508.
12. *Nutrition and Your Health: Dietary Guidelines for Americans.* 3rd ed. Washington, DC: US Dept of Agriculture and US Dept of Health and Human Services; 1990. Home and Garden Bulletin 232.
13. Curry S, Kristal A, Bowen D. An application of the stage model of behavior change to dietary fat reduction. *Health Educ Res.* 1992;7:97–105.
14. Zeigler R, Wilcox H, Mason T, Bill J, Virgo P. Seasonal variation in intake of carotenoids and vegetables and fruits among White men in New Jersey. *Am J Clin Nutr.* 1987;45:107–114.
15. Skinner CS, Strecher VJ, Hospers H. Physicians' recommendations for mammography: do tailored messages make a difference? *Am J Public Health.* 1994;84:43–49.

The Use of Vitamin Supplements and the Risk of Cataract among US Male Physicians

Johanna M. Seddon, MD, William G. Christen, DSc, JoAnn E. Manson, MD, Fran S. LaMotte, BS, Robert J. Glynn, DSc, Julie E. Buring, DSc, and Charles H. Hennekens, MD

ABSTRACT

Objectives. The purpose of this study was to examine prospectively the association between reported use of vitamin supplements and risk of cataract and cataract extraction.

Methods. The study population consisted of 17 744 participants in the Physicians' Health Study, a randomized trial of aspirin therapy and beta-carotene among US male physicians 40 to 84 years of age in 1982 who did not report cataract at baseline and provided complete information about vitamin supplementation and other risk factors for cataract. Self-reports of cataract and cataract extraction were confirmed by medical record review.

Results. During 60 months of follow-up, there were 370 incident cataracts and 109 cataract extractions. In comparison with physicians who did not use any supplements, those who took only multivitamins had a relative risk of cataract of 0.73 after adjustment for other risk factors. For cataract extraction, the corresponding relative risk was 0.79. Use of vitamin C and/or E supplements alone was not associated with a reduced risk of cataract, but the size of this subgroup was small.

Conclusions. These data suggest that men who took multivitamin supplements tended to experience a decreased risk of cataract and support the need for rigorous testing of this hypothesis in large-scale randomized trials in men and women. (*Am J Public Health.* 1994;84:788–792)

Introduction

Cataract is the leading cause of blindness and visual impairment in the world.[1] In the growing US elderly population, cataract is a leading cause of functional impairment and morbidity,[2] and cataract surgery has become the most common procedure performed among people more than 65 years of age.[2] Modifiable preventive factors such as diet (including vitamin supplementation) are receiving increased attention and merit careful evaluation.

Nutritional factors, particularly vitamins with antioxidant properties, may affect cataract development, possibly by protecting against free radical oxidative damage that results from many normal metabolic processes as well as exposure to ultraviolet light.[3,4] At present, data concerning the possible associations between vitamins and cataract are sparse and inconsistent.[5–14] We therefore sought to determine whether vitamin supplementation affected risk of cataract in a large prospective cohort of male physicians participating in the Physicians' Health Study.[15]

Methods

Study Population

The subjects and methods of the Physicians' Health Study have been described previously.[15] Briefly, the study is a randomized, double-masked, placebo-controlled trial that has tested (in a 2 × 2 factorial design) the effect of low-dose aspirin therapy (325 mg every other day) on the risks of cardiovascular disease and is continuing to evaluate beta-carotene (50 mg on alternate days) in the prevention of cancer and cardiovascular disease. The study involves 22 071 US male physicians who were 40 to 84 years of age in 1982. On January 25, 1988, the randomized aspirin component of the study was terminated early, principally as a result of the emergence of a statistically extreme 44% reduction in risk of a first myocardial infarction among those assigned to aspirin.

Supplement and Risk Factor Data

At baseline, participants completed questionnaires that requested information about dietary supplements. Questions included whether multiple vitamins had been taken regularly (never, past only, or current) and the number of years taken (if current). Questions were also asked about whether participants had ever regularly taken capsules containing only vitamin A, only vitamin C, and only vitamin E (never, past only, or current). For each vitamin currently used, partici-

Johanna M. Seddon and William G. Christen are with the Department of Ophthalmology, Harvard Medical School, Boston, Mass; Johanna M. Seddon is also with the Department of Epidemiology, Harvard School of Public Health. William G. Christen was and JoAnn E. Manson, Fran S. LaMotte, Robert J. Glynn, Julie E. Buring, and Charles H. Hennekens are with the Division of Preventive Medicine, Harvard Medical School and Brigham and Women's Hospital. Julie E. Buring and Charles H. Hennekens are also with the Department of Ambulatory Care and Prevention, Harvard Medical School.

Requests for reprints should be sent to Charles H. Hennekens, MD, Brigham and Women's Hospital, 900 Commonwealth Ave E, Boston, MA 02215-1204.

This paper was accepted August 10, 1993.

Editor's Note. See related editorial by Nestle (p 713) in this issue.

pants reported the number of years taken. In these analyses, subjects who reported current use of supplements were classified as supplement users. Since beta-carotene was an assigned treatment, those who reported vitamin A use were ineligible for enrollment in the Physicians' Health Study and were excluded. Although a complete dietary questionnaire was not administered, information was provided on a number of baseline characteristics, including cigarette smoking, alcohol use, history of diabetes mellitus, history of hypertension, history of high cholesterol, height, weight, physical activity, and parental history of myocardial infarction. This report includes 17 744 participants who provided complete information about vitamin supplements and other risk factors and reported that they did not have cataract at baseline.

Cataract Diagnoses

Follow-up questionnaires were sent at 6-month intervals during the first year and annually thereafter; these questionnaires requested information about compliance and occurrence of incident diseases, including cataracts. After report of a cataract diagnosis or extraction, a written consent requesting the name and address of the treating ophthalmologist was obtained. Ophthalmologists were contacted by mail and completed a cataract questionnaire that asked about the presence of lens opacities, visual acuity loss, cataract extraction, other ocular abnormalities that could explain visual acuity loss, cataract type, and primary cause (including age related, traumatic, congenital, inflammatory, or surgery or steroid induced). Alternatively, the ophthalmologists could supply photocopies of the relevant medical records.

The endpoints for this analysis were incident cataract and cataract extraction. Cataract was defined as follows: self-report, confirmed by medical record review, initially diagnosed after randomization, age related in origin, with best corrected visual acuity of 20/30 or worse and no alternate ocular pathology to explain the visual acuity loss. In the presence of alternate ocular pathology, a lens opacity was considered a cataract if, in the judgment of the ophthalmologist, the opacity was of sufficient severity to reduce visual acuity to 20/30 or worse when considered alone. Extraction was defined as the surgical removal of an incident cataract. For the present report, we followed the 17 744 eligible participants from randomization until January 25, 1988, the date the randomized aspirin component of the trial was terminated. The average length of follow-up was 60.2 months.

Data Analysis

Incidence rates of cataract and cataract extraction were obtained by dividing the number of incident cases by person-years of follow-up in each category of supplement use. Categories of vitamin use were vitamin C and/or E only, multivitamins only, multivitamins plus vitamin C and/or E, and no supplement intake. Relative risks (RRs) were computed as the rate of cataract in a specific category of supplement use divided by the corresponding rate among nonusers.

Data were analyzed separately for cataract diagnoses and extractions. Crude relative risk estimates controlled for age as well as randomized aspirin and beta-carotene assignment. In proportional hazard models used to control for multiple potential confounders, indicator terms were added for smoking (never, past, current), history of diabetes, history of hypertension (systolic blood pressure of 160 mm Hg or greater, diastolic blood pressure of 95 mm Hg or greater, or history of treatment for high blood pressure), history of high cholesterol (reported history of treatment for high cholesterol or reported blood cholesterol of 260 mg/dL or greater), obesity (body mass index of 27.8 kg/m^2 or greater), alcohol use (reported daily or weekly use), physical activity (reported vigorous exercise once per week or more), and parental history of myocardial infarction. Stratified analyses were performed to evaluate any possible modification of the effect of vitamin supplement use on cataract by category of smoking (never, past, current). Those not using supplements were the referent within each category of smoking.

For each relative risk, two-sided P values and 95% confidence intervals (CIs) were calculated.[16] Individuals, rather than eyes, were the unit of analysis because eyes were not examined independently and participants were classified according to the status of the worse eye. When the worse eye was excluded because of loss of visual acuity attributed to other ocular abnormalities or a cause that was not age related, the other eye was used for classification.

Results

There were 370 incident cataracts and 109 cataract extractions during follow-up. At baseline, 12.5% of the participants reported multivitamin use only, 7.5% used multivitamin supplements plus C and/or E, and 4.1% reported vitamin C and/or E supplementation. The distribution of baseline characteristics that are possible risk factors for cataract are presented according to supplement status in Table 1. Vitamin supplement users tended to be older and reported more physical activity, more hypertension, and less obesity.

As seen in Table 2, in analyses adjusted for age and randomized treatment assignment, multivitamin-only users had a marginally significant decreased risk of incident cataract relative to nonsupplementers (RR = 0.75, 95% CI = 0.55, 1.01). After adjustment for other potential cataract risk factors, the relative risk was 0.73 (95% CI = 0.54, 0.99). Users of vitamin C and/or E alone, however, did not appear to have a reduced risk of cataract, but the numbers were small. Multivitamin users who also used supplements containing only vitamin C and/or E had a risk of incident cataract similar to users of multivitamins only. The relative risk of cataract decreased with increasing duration of any use of multivitamins, an apparent trend of borderline statistical significance (P = .06) (Table 3).

The adjusted relative risk for cataract extraction among multivitamin users only (0.79, 95% CI = 0.46, 1.35) was similar to that for cataract diagnosis. Risk of cataract extraction was not reduced among participants taking vitamin C and/or E alone. However, as with cataract diagnosis, the number of cataract extractions in this subgroup was small.

The association between any multivitamin use and risk of cataract within categories of cigarette smoking was also assessed. Among those who never smoked, the relative risk was 1.10 (95% CI = 0.76, 1.60) after adjustment for other risk factors. However, among past and current smokers, the risk of cataract was reduced among multivitamin users relative to nonusers (RR = 0.61, 95% CI = 0.42, 0.90, and RR = 0.38, 95% CI = 0.16, 0.92, respectively). This trend for a reduction in risk among multivitamin users who smoked was statistically significant (P = .008). Numbers were too small to evaluate smoking among users of vitamins C and/or E only or the endpoint cataract extraction.

TABLE 1—Differences in Sample's Baseline Characteristics That Are Known or Potential Risk Factors for Cataract, by Supplement Status

	Vitamin C and/or E Only	Multivitamin Only	Multivitamin and C and/or E	No Vitamins
Total n	729	2222	1331	13 462
Mean age, y	54.0	55.4	54.5	52.2
Diabetes, %	0.8	3.7	3.2	2.1
Physical activity, %[a]	76.0	74.0	78.0	71.6
History of high cholesterol, %[b]	5.9	7.7	8.3	6.5
History of parental myocardial infarction, %	10.7	12.7	12.5	13.7
Hypertension, %[c]	16.3	17.6	14.5	13.2
Obesity, %[d]	10.0	12.4	11.3	14.0
Smoking, %				
Never	53.5	45.4	49.6	51.6
Past	39.9	42.2	40.7	38.1
Current	6.6	12.3	9.7	10.3
Alcohol, %				
Daily	24.3	28.9	27.1	23.4
Weekly	49.9	42.8	44.6	50.5

[a]Defined as reported vigorous exercise once per week or more.
[b]Defined as reported high cholesterol or reported blood cholesterol of 260 mg/dL or greater.
[c]Defined as reported systolic blood pressure of 160 mm Hg or greater, diastolic blood pressure of 95 mm Hg or greater, or history of treatment for high blood pressure.
[d]Defined as body mass index of 27.8 kg/m² or greater.

TABLE 2—Association between Vitamin Supplement Use and Risk of Cataract Diagnosis, by Supplement Status

	Vitamin C and/or E Only	Multivitamin Only	Multivitamin and C and/or E	No Vitamins
No. of cataracts	22	51	29	268
Total person-years of observation	3598	10 983	6584	67 010
Short model[a]				
Relative risk	1.26	0.75	0.77	1.0
95% confidence interval	0.81, 1.94	0.55, 1.01	0.52, 1.13	. . .
P	.30	.06	.18	. . .
Full model[b]				
Relative risk	1.32	0.73	0.77	1.0
95% confidence interval	0.85, 2.04	0.54, 0.99	0.53, 1.14	. . .
P	.21	.04	.19	. . .

[a]Adjusted for age and randomized treatment assignment.
[b]Adjusted for age, randomized treatment assignment, history of diabetes mellitus, history of hypertension (systolic blood pressure of 160 mm Hg or greater, diastolic blood pressure of 95 mm Hg or greater, or history of treatment of high blood pressure), obesity (body mass index of 27.8 kg/m² or greater), alcohol (reported daily or weekly use), physical activity (reported vigorous exercise once per week or more), parental history of myocardial infarction, high cholesterol (reported treatment for high cholesterol or reported blood cholesterol of 260 mg/dL or greater), and smoking status (never, past, or current).

Discussion

These prospective data are compatible with the possibility that vitamin users may have a lower risk of cataract and cataract extraction. If real, even modest decreases in risk such as those suggested in this study are potentially important, given the high prevalence of cataract, the enormous costs associated with cataract extraction, and the feasibility of dietary changes or micronutrient supplementation.

The prospective design of this study precludes the possibility of bias in reporting supplementation use based on disease outcome, since vitamin supplement use was reported at baseline prior to questions regarding the cataract outcomes and only incident cataracts diagnosed after this baseline period were included in the analyses. Because of the mail questionnaire design, the diagnoses of cataract and cataract extraction were based on medical record confirmation. There were essentially no false-positive cases of cataract extraction and, because we required medical record documentation, probably very few if any false-positive incident cataracts. Although there may have been cases that were not ascertained (false negatives), such underascertainment should not bias the relative risks in a cohort study if not associated with exposure.[17,18] Differential misclassification of cataract based on supplement use was extremely unlikely, since medical records were reviewed without knowledge of the participants' exposure data and treating ophthalmologists would have been generally unaware of the participants' dietary habits or the vitamin–cataract hypothesis, which has only been reported recently.

It is possible that multivitamin supplement users may be more health conscious and may, therefore, have more (or less) medical contacts, which would increase (or decrease) the likelihood of having a cataract diagnosed. However, all participants in this study are physicians, so their overall level of medical care is likely to be reasonably uniform, with no large differences between users and nonusers of supplements. Misclassification of vitamin intake is unlikely, since use of supplements is usually accurately recalled.[19] Such misclassification, if present, would probably be random and would tend to bias associations toward the null. Incomplete follow-up is not an issue in this study. Morbidity follow-up was 99.3% complete as of January 25, 1988, and ocular medical records were obtained for more than 90% of both users and nonusers of supplements.

Several known or potential risk factors for cataract, such as diabetes and smoking status, were controlled in the analyses. Of course, the effects of unmeasured or unknown factors cannot be controlled in an observational study. It is possible that users of multivitamins differed from nonusers in ways that were not measured and therefore, could not be adjusted for in these analyses. For example, we were unable to control accurately for the effects of sun exposure. It is unlikely that there are large differences in

sun exposure, however, among this homogeneous group of male physicians, with most having no occupational exposure to sunlight. This possibility is supported by the lack of an association between geographic region and cataract when this variable was included in the multivariate models.

Foods rich in vitamins with antioxidant properties are now considered to have the potential to prevent many chronic diseases, including coronary heart disease and cancer.[20] A similar role has been hypothesized in the formation and progression of cataract[21] that centers on the protective action of certain micronutrients against free radical oxidative damage.[22,23] Experimental studies have demonstrated that both vitamin C[22] and vitamin E[23] prevent lipid peroxidation in response to photo-oxidative assault and limit lens damage after oxidative insult in animals fed these vitamins.[24,25] A few epidemiologic studies have also evaluated the association between antioxidant nutrient status and cataract.[5–14] Some of these cross-sectional or retrospective studies have found an inverse association between cataract and plasma[5] or dietary[7,8] vitamin C, plasma[14] or dietary[7,8] vitamin E, and use of multivitamins.[8] In contrast, others have not shown an inverse association with vitamins C, E, or beta-carotene[9,10,14] or with use of multivitamins.[7] Data from two studies that used prospectively identified cases have also been conflicting. Inverse associations with cataract or cataract extraction were reported for dietary carotenoids,[12] long-term vitamin C supplementation,[12] and plasma levels of vitamin E and beta-carotene.[13] On the other hand, no associations were observed between the cataract endpoint and dietary vitamins C or E.[12]

Our prospective data, combined with evidence from previous studies, are compatible with the possibility that multivitamin supplements are related to cataract. Although the numbers are small, our findings regarding separate supplementation with vitamins C and E, which both have antioxidant properties, do not suggest a protective effect for these vitamins used either alone or in addition to multivitamins. The antioxidant properties of multivitamins, which contain vitamins C and E, represent a plausible mechanism by which risk of cataract may be reduced. However, there may be other components of multivitamins separate from antioxidants that offer protection. Further laboratory and clinical research is needed to elucidate mechanisms to understand how multivitamin supplementation may affect risk of cataract.

TABLE 3—Association between Duration of Multivitamin Supplement Use and Risk of Cataract

	Duration of Use, y[a]			
	<5	5–10	>10	*P* for Trend
No. of cataracts	32	11	18	...
No. of person-years	7847	2598	3229	...
Short model[b]				
Relative risk	0.88	0.75	0.69	.07
95% confidence interval	0.61, 1.27	0.41, 1.36	0.43, 1.11	...
Full model[c]				
Relative risk	0.88	0.74	0.68	.06
95% confidence interval	0.61, 1.27	0.40, 1.35	0.42, 1.10	...

[a] n = 3553 (2222 multivitamin users and 1331 multivitamin plus C and/or E users).
[b] Adjusted for age and randomized treatment assignment.
[c] Adjusted for age, randomized treatment assignment, history of diabetes mellitus, history of hypertension (systolic blood pressure of 160 mm Hg or greater, diastolic blood pressure of 95 mm Hg or greater, or history of treatment of high blood pressure), obesity (body mass index of 27.8 kg/m^2 or greater), alcohol (reported daily or weekly use), physical activity (reported vigorous exercise once per week or more), parental history of myocardial infarction, high cholesterol (reported high cholesterol or reported blood cholesterol of 260 mg/dL or greater), and smoking status (never, past, or current).

Results of our analyses also raise the possibility of a benefit of multivitamin supplementation on risk of cataract, primarily among the subgroups of past and current smokers. Although it is plausible that supplementation might help particularly those with reduced levels of nutrients related to oxidative stress, as may occur among smokers, our subgroup findings are speculative, based on small numbers, and require further study.

As is the case with detection of small to moderate benefits, observational studies have amounts of uncontrolled or uncontrollable confounding that may be as large as the effects being postulated. Thus, the only reliable means to determine whether antioxidant vitamins decrease cataract is through large-scale randomized trials.[17,26] In this regard, the National Eye Institute's Age-Related Eye Disease Study[26] is testing vitamins C, E, and beta-carotene, as well as zinc; the Physicians' Health Study[27] is evaluating beta-carotene; and the Women's Health Study[28] is testing beta-carotene and vitamin E. All of these data will provide relevant and complementary information to a body of evidence on which reliable clinical decision making and public policy can safely be based.

In summary, these prospective data from the Physicians' Health Study suggest that multivitamin supplementation may be associated with a modest decrease in the risk of cataract. Results were marginally significant and were derived from an observational study design. Findings support the need for large-scale randomized trials in men and women to evaluate definitively the possibility that micronutrient supplementation may reduce the worldwide burden of blindness due to cataract. □

Acknowledgment

This research was supported by National Institutes of Health grant EY03355.

References

1. Kupfer C. Six main causes of blindness. In: Wilson J, ed. *World Blindness and Its Prevention.* New York, NY: Oxford University Press; 1984.
2. Stark WJ, Sommer A, Smith RE. Changing trends in intraocular lens implantation. *Arch Ophthalmol.* 1989;107:1441–1444.
3. Taylor HR, West SK, Rosenthal FS, et al. Effect of ultraviolet radiation on cataract formation. *N Engl J Med.* 1988;319:1429–1433.
4. Hiller R, Giacometti L, Yuen K. Sunlight and cataract: an epidemiologic investigation. *Am J Epidemiol.* 1977;105:450–459.
5. Jacques PF, Hartz SC, Chylack LT, McGandy RB, Sadowski JA. Nutritional status in persons with and without senile cataract: blood vitamin and mineral levels. *Am J Clin Nutr.* 1988;48:152–158.
6. Jacques PF, Chylack LT. Epidemiologic evidence of a role for the antioxidant vitamins and carotenoids in cataract prevention. *Am J Clin Nutr.* 1991;53:352S–355S.
7. Robertson JM, Donner AP, Trevithick JR. Vitamin E intake and risk of cataracts in

humans. *Ann NY Acad Sci.* 1989;570:372–382.
8. Leske MC, Chylack LT, Wu S-Y. The lens opacities case-control study: risk factors for cataract. *Arch Ophthalmol.* 1991;109:244–251.
9. Italian-American Cataract Study Group. Risk factors for age-related cortical, nuclear, and posterior subcapsular cataracts. *Am J Epidemiol.* 1991;133:541–553.
10. Goldberg J, Flowerdew G, Tso MOM, Brody JA. Age-related macular degeneration and cataract: are dietary antioxidants protective? *Am J Epidemiol.* 1988;128:904–905.
11. Mohan M, Sperduto RD, Angra SK, et al. India-US case-control study of age-related cataracts. *Arch Ophthalmol.* 1989;107:670–676.
12. Hankinson SE, Stampfer MJ, Seddon JM, et al. Nutrient intake and cataract extraction in women: a prospective study. *Br Med J.* 1992;305:335–339.
13. Knekt P, Heliovaara M, Rissanen A, Aromaa A, Aaran R-K. Serum antioxidant vitamins and risk of cataract. *Br Med J.* 1992;305:1392–1394.
14. Vitale S, West S, Hallfrisch J, et al. Plasma antioxidants and risk of cortical and nuclear cataract. *Epidemiology.* 1993;4:195–203.
15. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med.* 1989;321:129–135.
16. Miettinen O. Estimability and estimation in case-referent studies. *Am J Epidemiol.* 1976;103:226–235.
17. Hennekens CH, Buring JE. *Epidemiology in Medicine.* Boston, Mass: Little, Brown; 1987.
18. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol.* 1977;105:488–495.
19. Willett WC. *Nutritional Epidemiology.* New York, NY: Oxford University Press; 1989.
20. Diplock AT. Antioxidant nutrients and disease prevention: an overview. *Am J Clin Nutr.* 1991;53:189S–193S.
21. Bunce GE, Kinoshita J, Horwitz J. Nutritional factors in cataract. *Annu Rev Nutr.* 1990;10:233–254.
22. Varma SD, Srivastava VK, Richards DD. Photoperoxidation in lens and cataract formation: preventive role of superoxide dismutase, catalase and vitamin C. *Ophthalmic Res.* 1982;14:167–175.
23. Varma SD, Beachy NA, Richards DD. Photoperoxidation of lens lipids: prevention by vitamin E. *Photochem Photobiol.* 1982;36:623–626.
24. Blondin J, Baragi VK, Schwartz ER, Sadowski J, Taylor A. Prevention of eye lens protein damage by dietary vitamin C. *Invest Ophthalmol Vis Sci.* 1986;27:9–14.
25. Bhuyan KC, Bhuyan DK, Podos SM. The role of vitamin E in therapy of cataract in animals. *Ann NY Acad Sci.* 1982;393:169–171.
26. Sperduto RD, Ferris FL, Kurinij N. Do we have a nutritional treatment for age-related cataract or macular degeneration? *Arch Ophthalmol.* 1990;108:1403–1405.
27. Hennekens CH, Eberlein K. A randomized trial of aspirin and beta-carotene among US physicians. *Prev Med.* 1985;14:165–168.
28. Buring JE, Hennekens CH. The Women's Health Study: rationale and background. *J Myocardial Ischemia.* 1992;4:30–40.

High Temporal, Geographic, and Income Variation in Body Mass Index among Adults in Brazil

Rosely Sichieri, MD, PhD, Denise C. Coitinho, MPH, Marília M. Leão, Elisabetta Recine, MPH, and James E. Everhart, MD, MPH

ABSTRACT

Objectives. Population-based data on body mass index for developing countries are scarce. Body mass index data from two Brazilian surveys were examined to determine regional and temporal variations in the prevalences of underweight, overweight, and obesity.

Methods. Nationwide surveys in 1974/75 and 1989 collected anthropometric data in Brazil from 55 000 and 14 455 households, respectively. Trained interviewers used the same methods to measure weight and stature in both surveys, and survey designs were identical. Prevalences of underweight, overweight, and obesity were determined for persons 18 years of age and older.

Results. In the 1989 survey, body mass index varied greatly according to region of the country, urbanization, and income. In the wealthier South, the prevalence of overweight/obesity was the highest and the prevalence of underweight was the lowest; in the poorer rural Northeast, these patterns were reversed. For both surveys, overweight/obesity was more common among women than among men and peaked at age 45 to 64 years in both sexes. Over the 15 years between surveys, the prevalence of both overweight and obesity increased strikingly.

Conclusions. In contrast to findings in developed countries, obesity in Brazil was positively associated with income and was much more prevalent among women than among men. For Brazilian women, the overall prevalence of overweight was nearly as high as that among women in the United States. (*Am J Public Health.* 1994;84:793–798)

Introduction

Obesity is recognized as a condition characteristic of high-income societies.[1–3] Little is known, however, about the pattern of weight change within populations of developing countries. Increased mortality rates owing to obesity-related chronic diseases indicate that obesity may be an important risk factor in these countries,[4,5] but population-based data are scarce. In 1974/75 and 1989, the Brazilian Institute of Geography and Statistics conducted national representative surveys that measured the weight and stature of the participants. The purpose of this report is to describe demographic variations in the prevalence of underweight, overweight, and obesity based on body mass index and to compare the increased prevalence of overweight and obesity over the 15 years between surveys with such prevalences in developed countries.

Population and Methods

Study Population

Stature and weight have been measured in two nationwide nutritional surveys in Brazil: the Estudo Nacional da Despesa Familiar,[6] conducted in 1974/75, and the Pesquisa Nacional de Saúde e Nutrição,[7] conducted in 1989. The first was based on a national probability sample of 55 000 Brazilian households; the second was based on a representative population sample of 14 455 households. For both surveys, samples were selected from the urban and rural districts of the five major geographical regions of the country; the only area not included was the sparsely populated rural North. For both surveys, participants' households were selected using two-stage probability sampling. In the first stage, primary sampling units were selected from the nine major urban/rural regions of the country. In the second stage, households were sampled from the primary sampling units. Probability of selection of primary sampling units and households was proportional to the population in the nine areas. Household nonresponse for the 1989 survey was 10.8%, of which 99% was owing to failure to locate household occupants for interview and examination.

Measurement of the Variables

Stature and weight were collected in the households by two trained interviewers, with the interviewee wearing light clothes and no shoes. The scales were calibrated at each household, and for each person the measurements were taken twice. Scales were precise to 100 g. Height was measured using a platform with an attached measuring bar, precision was within 0.1 cm.

Per capita income was calculated as the preceding month's family income divided by the number of persons living on

Rosely Sichieri is with the Centro de Ciências Biológicas e da Saúde Universidade Estadual de Maringá, Paraná, Brazil. Denise C. Coitinho, Marília M. Leão, and Elisabetta Recine are with the Centro Nacional de Epidemiologia e Informação–CENEI, Sarah–Hospital de Medicina do Aparelho Locomotor, Brasília, Brazil. James E. Everhart is with the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Md.

Requests for reprints should be sent to Rosely Sichieri, MD, PhD, Instituto de Medicina Social, UERJ, Rua S. Francisco Xavier 524, 7o andar, Bloco E, Rio de Janeiro, CEP 20550-012 Brasil.

This paper was accepted October 21, 1993.

TABLE 1—Sample Size and Prevalence for Underweight, Overweight, and Obesity[a] among Brazilian Adults Aged 18 and Older in Two National Surveys, by Sex

	Pesquisa Nacional de Saúde e Nutrição, 1989				Estudo Nacional da Despesa Familiar, 1974/75			
	Sample Size		Prevalence, %		Sample Size		Prevalence, %	
	No.	%	Men (n = 16 783)	Women (n = 17 168)	No.	%	Men (n = 63 138)	Women (n = 65 169)
Underweight (BMI <20)	5 432	16.0	15.4	16.8	32 318	25.2	24.1	26.3
Normal weight (BMI 20–24)	17 383	51.2	57.1	45.7	68 546	53.4	58.9	47.9
Overweight (BMI 25–29)	8 318	24.5	22.5	26.4	21 416	16.7	14.4	18.9
Obesity (BMI >30)	2 818	8.3	4.8	11.7	6 027	4.7	2.5	6.9

[a]Based on body mass index (BMI), which is defined as weight in kilograms divided by height in square meters (kg/m²).

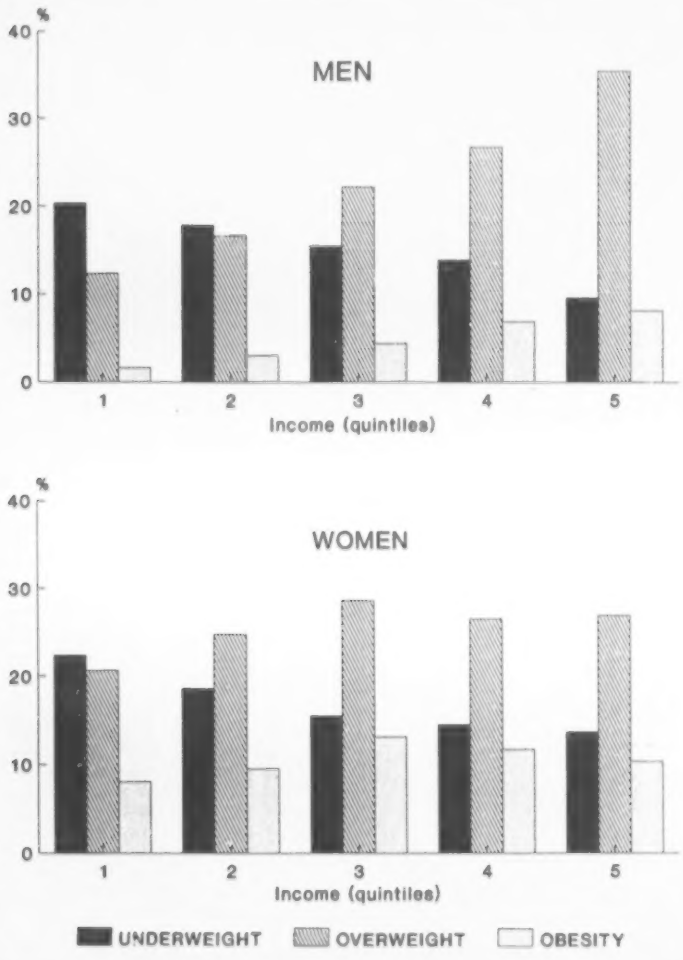


FIGURE 1—Age-adjusted prevalence (%) of underweight (body mass index [BMI] < 20 kg/m²) and overweight/obesity (BMI ≥ 25 kg/m²), by income and sex: Brazil, 1989.

that budget. Family income was categorized into quintiles and also as being above or below twice the Brazilian minimum wage per month (US $150 in July 1989), the estimated income necessary for the survival of a family of four.

Analytical Procedures

Body mass index, defined as weight in kilograms divided by height in square meters (kg/m²), was classified according to standards of the World Health Organization (WHO)[8] with a modification in terminology. Underweight was defined as a body mass index of less than 20 kg/m², normal weight as 20 to 24 kg/m², overweight as 25 to 30 kg/m² (WHO obesity grade 1), and obesity as above 30 kg/m² (WHO obesity grades 2 and 3). The overweight and obesity categories were combined since findings for these two categories were very similar. Persons missing data were excluded; there were 115 missing values for body mass index and 800 missing values for income in the 1989 survey.

The SUDAAN DESCRIPT procedure[9] was used to estimate the population percentages of underweight, overweight, and obesity, age-standardized by 10-year age groups to the estimated Brazilian population in 1989. Logistic regression with RTILOGIT[10] was used to examine the age, sex, and income-adjusted associations of underweight and overweight/obesity with urban or rural residence and geographic region of the country. These regions were entered in the models as four indicator variables, with the Southeast being the reference region.

Brazilian prevalence of overweight/obesity (body mass index ≥ 25 kg/m²) was compared with prevalences for Great Britain in 1980,[11] Canada in 1981,[11] and United States from 1976 to 1980.[12]

Results

Only 45.7% of women and 57.1% of men were classified as having normal weight in the 1989 survey. In the 1974/75 survey, underweight was more prevalent than overweight/obesity, but by 1989 the reverse was the case (Table 1).

Prevalence of Underweight

The overall prevalence of underweight was 16.0%, with 13 million Brazilian adults classified as underweight in 1989. Overall prevalence was slightly higher among women than among men (Table 1). For both sexes, the age-adjusted prevalence was higher in rural than in urban areas (Table 2). The highest prevalence of underweight was found in the Northeast and the lowest was in the South. For both sexes, the prevalence of underweight decreased with increasing income (Figure 1).

In the multivariate logistic regression analysis, age, income, sex, rural residence, and geographic region were significant independent risk factors for underweight (Table 3). Living in the South, living in a urban area, and being relatively affluent decreased the odds of being underweight.

Prevalence of Overweight and Obesity

The overall prevalence of overweight/obesity in 1989 was 32.8%. Women had a greater prevalence of overweight/obesity than men (Table 1). Among men, the prevalence was much higher in urban than in rural areas, whereas among women, a clearly higher urban prevalence occurred only in the Northeast and the Southeast (Table 2). Among the regions, the highest prevalence of overweight/obesity was found in the South and the lowest was in the Northeast.

Prevalence of overweight/obesity increased with income among men but did not increase beyond the third quintile of income among women (Figure 1). In the logistic regression analysis, all variables studied were significantly associated with overweight/obesity, with associations opposite to those found with underweight (Table 3). The only exception was that women were at higher risk than men of both underweight and overweight/obesity.

Trends in Body Mass Index in Brazil

The trend for the 15-year period studied was a striking reduction in the prevalence of underweight and an increase in the prevalence of overweight/obesity (Figure 2). Decrease in the prevalence of underweight was about 36% for both sexes combined. For men, overweight increased by 56.3% and obesity increased by 92.0% whereas for women, the increases were less pronounced: 39.7% for overweight and 69.6% for obesity (Table 1).

TABLE 2—Age-Adjusted Prevalence (%) of Underweight and Overweight/Obesity, by Regions of the Country, Urban or Rural Residence, and Sex: Brazil, 1989

Regions	Men		Women	
	Urban (n = 9291)	Rural (n = 7492)	Urban (n = 10 595)	Rural (n = 6573)
Underweight (BMI < 20 kg/m²)[a]				
North	11.9	. . .	16.8	. . .
Northeast	16.5	22.7	19.1	29.0
Southeast	14.5	21.3	14.5	17.2
South	9.1	12.0	11.4	13.8
Midwest	13.4	18.4	17.8	20.9
Overweight/obesity (BMI ≥ 25 kg/m²)[a]				
North	32.0	. . .	37.3	. . .
Northeast	25.4	10.3	33.8	20.7
Southeast	31.6	16.8	41.8	34.7
South	38.2	24.7	42.3	41.9
Midwest	31.5	17.9	36.2	34.2

[a]BMI = body mass index.

TABLE 3—Odds Ratios[a] and 95% Confidence Intervals for Underweight and Overweight/Obesity, According to Age, Income, Sex, Urban or Rural Residence, and the Five Geographic Regions of the Country: Brazil, 1989

	Underweight (BMI < 20 kg/m²)[b]		Overweight/Obesity (BMI ≥ 25 kg/m²)[b]	
	Odds Ratio	95% Confidence Interval	Odds Ratio	95% Confidence Interval
Age, y[c]				
20	1.87	1.12, 3.08	0.21	0.13, 0.33
40	1.03	0.85, 1.25	0.74	0.62, 0.88
50	1.00	. . .	1.00	. . .
60	1.15	0.94, 1.42	1.08	0.89, 1.31
80	2.61	1.31, 5.19	0.65	0.34, 1.23
Income (US $100)	0.91	0.88, 0.94	1.07	1.05, 1.09
Sex (women/men)	1.38	1.29, 1.46	1.72	1.63, 1.80
Residence (urban/rural)	0.82	0.77, 0.89	1.60	1.49, 1.71
Region				
North	0.92	0.76, 1.12	0.92	0.80, 1.07
Northeast	1.13	1.05, 1.22	0.71	0.67, 0.75
Southeast	1.00	. . .	1.00	. . .
South	0.67	0.61, 0.75	1.18	1.10, 1.27
Midwest	1.02	0.90, 1.16	0.91	0.82, 1.01

[a]Odds ratio for a single variable adjusted for all the other variables in the model and calculated by multiple logistic regression.
[b]BMI = body mass index.
[c]Model adjusted for age and age squared.

International Comparisons

The overall prevalence of overweight/obesity among men in Brazil was much lower than that in some developed countries (Figure 3). However, when the analysis was restricted to Brazilian men

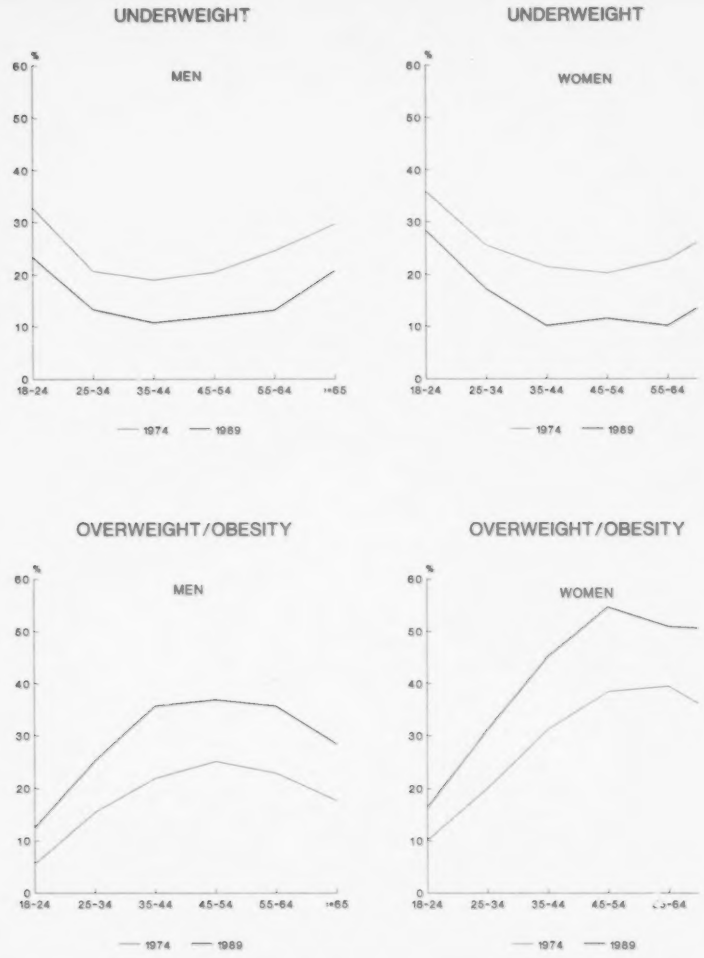


FIGURE 2—Prevalence (%) of underweight (body mass index < 20 kg/m²) and overweight/obesity (BMI ≥ 25 kg/m²) in two surveys—Estudo Nacional de Despesa Familiar, 1974/75, and Pesquisa Nacional de Saúde e Nutrição, 1989—by sex.

with family income higher than twice the Brazilian minimum wage, the prevalence of overweight/obesity was comparable to that in the other countries. Among women, the prevalence of overweight/obesity, both overall as well as among those with higher income, was comparable to that among US women and higher than that in the other countries.

Discussion

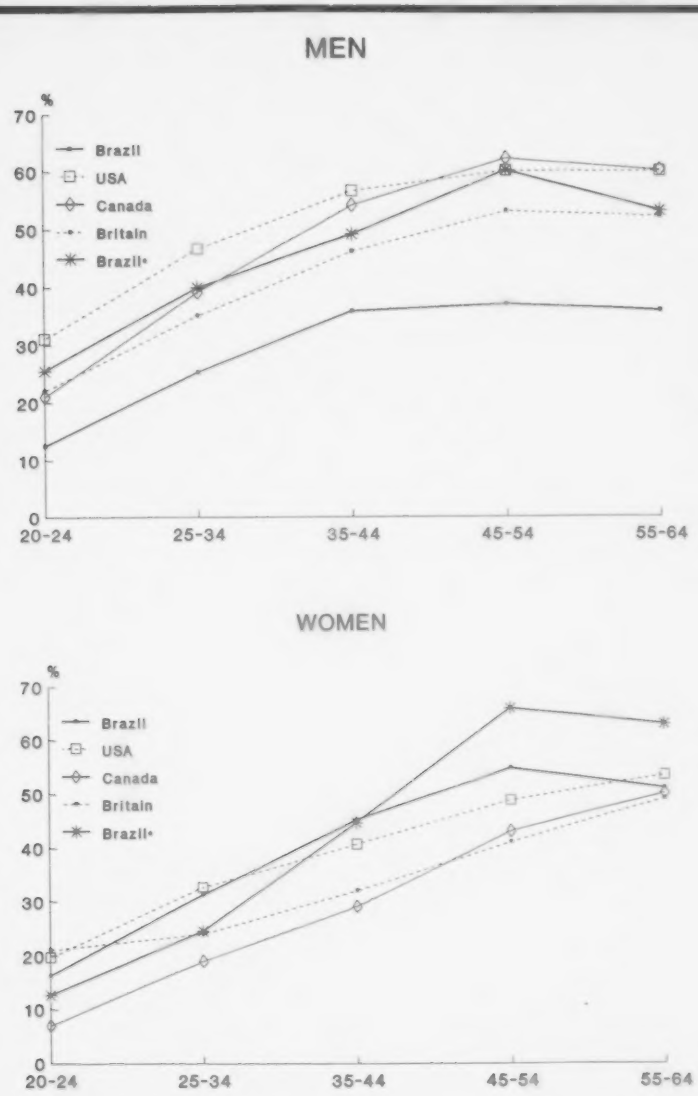
There was a striking difference in body mass index in Brazil according to sex, region, urbanization, and income. Women were almost 2.5 times as likely as men to be obese, a degree of gender difference unobserved in developed countries (Figure 3). However, the pattern in Brazil may be changing toward that of these other countries. The prevalence of overweight/obesity increased more rapidly for men than for women between the two surveys, and the greatest increase was among young men. Also, at the highest income levels, the prevalence of overweight was more common among men than among women. These shifts toward the attributes of more developed countries raise the possibility that Brazil is currently in the midst of changes that occurred among more economically developed countries before population-based surveys were in general use. A stepwise increase of overweight/obesity with income was observed for men, whereas other studies have found weakly positive associations.[13,14] In contrast, a negative association between socioeconomic status and body mass index has been reported in many studies of women.[15–17] Among Brazilian women, the prevalence of overweight/obesity was highest in the third quintile of income and slightly lower for the two higher quintiles.

Both urbanization and the more developed geographic regions in Brazil were positively associated with the prevalence of overweight/obesity and negatively associated with the prevalence of underweight, independently of income (Table 3). Additional adjustment for race did not explain these urban/rural and regional differences. Comparative data on race are not available for the 1974/75 survey.

The striking variations in body mass index by urbanization, income, and region of the country indicate that environmental factors are important determinants of obesity. Both the available supply of goods and services in urban areas and the nutritional status of the urban populations are better, as evidenced by greater mean heights among urban residents.[18] Chronic malnutrition and acute diarrheal illness in areas of poorer sanitation may also have contributed to regional and urban/rural differences in underweight.

The marked increase in the prevalence of obesity in a short space of time, as shown in this study, has not been observed in developed countries.[19–21] Our analysis supports the concept proposed by Price et al.[21] in 1991 that a general improvement in environmental conditions that predisposed all individuals to obesity would shift the overall body mass index distribution to higher values. However, the recent secular trends in developed countries have shown only an increase in the overweight component of the body mass index distribution without a substantial increase in the mean or median.[19,21]

Measures of stature and weight were carefully performed in both surveys; thus, the differences in body mass index between surveys are unlikely to be a result of inaccurate anthropometric measures. Moreover, any misclassification of body mass index would have been expected in both surveys since the measurement of weight and stature in both surveys was



Note. Brazilian curves marked with stars excluded the population below the poverty level (two minimum wages per month per family).

FIGURE 3—Prevalence (%) of overweight/obesity (body mass index ≥ 25 kg/m²) for Great Britain in 1980,[11] Canada in 1981,[11] United States from 1976 to 1980,[12] and Brazil in 1989.

very similar. Comparisons among countries also should not depend on methodological differences between surveys since all data were based on carefully measured weight and stature.

Differences in energy intake and physical activity are the obvious explanation for Brazil's rapid changes and geographic variations in the prevalence of overweight and obesity. Dietary data collected in the 1974/75 survey show a positive relationship of mean energy intake with both body weight and income.[22] Thus, part of the increase in body mass index in Brazil may be related to increased energy intake. However, even when controlling for income, a correlate of energy intake, marked urban/rural differences in body mass index persist. Perhaps greater physical activity is a reason for the lower risk of overweight and obesity in rural areas where manual labor remains common.

The rapid increase and wide demographic variation in the prevalence of overweight/obesity in Brazil present opportunities for studying the development of this condition. Such opportunities are not available in countries with slower anthropometric changes and less distinct demographic variations. □

References

1. Bray GA. Obesity: basic considerations and clinical approaches. *Dis Mon.* 1989;35: 449–537.
2. Jalkanen L, Tuomilehto J, Nissinen A, Puska P. Changes in body mass index in a Finnish population. *J Intern Med.* 1989;226: 163–170.
3. Kuskowska-Wolk A, Rossner S. Prevalence of obesity in Sweden: cross-sectional study of a representative adult population. *J Intern Med.* 1990;227:241–246.
4. Dowd JE, Manton KG. Forecasting chronic disease risk factors in developing countries. *Int J Epidemiol.* 1990;19:1019–1036.
5. Becker RA, Lima DD, Lima JTF, Costa ML Jr. *Investigações sobre Perfis de Saúde: Brasil, 1984.* Brasília, Brasil: Centro de Documentação do Ministério da Saude; 1989.
6. *Metodologia do Estudo Nacional da Despesa Familiar. Objetivos, Descrição e Metodologia Usada no ENDEF.* Rio de Janeiro, Brasil: Fundação Instituto Brasileiro de Geografia e Estatística; November 1983.
7. *Pesquisa Nacional sobre Saúde e Nutrição. PNSN—1989. Arquivo de Dados da Pesquisa.* Brasília, Brasil: Instituto Nacional de Alimentação e Nutrição; April 1990.
8. *Diet, Nutrition and Prevention of Chronic Disease.* Report of a WHO study group. Geneva, Switzerland: World Health Organization; 1990. Technical Report Series 797.
9. *Software for Survey Data Analysis (SUDAAN), Version 5.40,* Research Triangle, NC: Research Triangle Institute; 1991.
10. Shah BV, Folsom RE, Harrell FE, et al. *Survey Data Analysis Software for Logistic Regression.* Research Triangle, NC: Research Triangle Institute; 1984.
11. Millar WJ, Stephens T. The prevalence of overweight and obesity in Britain, Canada, and United States. *Am J Public Health.* 1987;77:38–41.
12. Najjar MF, Rowland M. Anthropometric reference data and prevalence of overweight. *Vital Health Stat [11].* 1987;238. DHHS publication PHS 87–1618.
13. Flegal KM, Harlan WR, Landis JR. Secular trends in body mass index and skinfold thickness with socioeconomic factors in young adult men. *Am J Clin Nutr.* 1988;48: 544–551.
14. Fehily AM, Phillips KM, Yarnell JWG. Diet, smoking, social class, and body mass index in the Caerphilly Heart Disease Study. *Am J Clin Nutr.* 1984;40:827–833.

15. Flegal KM, Harlan WR, Landis JR. Secular trends in body mass index and skinfold thickness with socioeconomic factors in young adult women. *Am J Clin Nutr.* 1988;48:535–543.
16. Goldblatt PB, Moore ME, Stunkard AJ. Social factors in obesity. *JAMA.* 1965;192: 1039–1044.
17. Rimm I, Rimm A. Association between socioeconomic status and obesity in 59 556 women. *Prev Med.* 1974;3:543–572.
18. Eveleth PB. Population differences in growth. In: Tanner JM, Falkner FT, eds. *Human Growth: A Comprehensive Treatise.* New York, NY: Plenum Press; 1986:221–238.
19. Harlan WR, Landis JR, Flegal KM, Davis CS, Miller ME. Secular trends in body mass index in the United States, 1960–1980. *Am J Epidemiol.* 1988;128:1065–1074.
20. Rissanen A, Heliovaara M, Aromaa A. Overweight and anthropometric changes in adulthood: a prospective study of 17 000 Finns. *Int J Obes.* 1988;12:391–401.
21. Price RA, Ness R, Sorensen TIA. Changes in commingled body mass index distributions associated with secular trends in overweight among Danish young men. *Am J Epidemiol.* 1991;133:501–510.
22. *Estudo Nacional da Despesa Familiar—ENDEF; Consumo Alimentar-Antropometria.* Rio de Janeiro, Brazil: Fundação Instituto Brasileiro de Geografia e Estatística; 1977.

Call for Abstracts for Second International Conference on Dietary Assessment Methods

The Second International Conference on Dietary Assessment Methods will be held in Boston, Mass, January 22 through 24, 1995. The conference is designed to facilitate ongoing exchange of information, stimulate national and international collaborative research, and encourage innovative approaches to improving methods for collecting and analyzing dietary data. This international exchange of knowledge and ideas will enhance our understanding of the role of diet in the etiology and prevention of disease and will facilitate the development of nutrition programs and policies leading to improved world health.

Participants will come from academic and health care settings, food industries, and government agencies around the world. Speakers and faculty will include a broad spectrum of internationally recognized experts in dietary assessment methodologies. This conference will provide opportunities to discuss subjects that include the cognitive influence in dietary assessment methodology, experiences from diverse cultures, food consumption patterns, biomarkers of dietary intake, calibration/validation methodology, statistical interpretation and adjustment of dietary data, surveillance and monitoring, and food composition/contaminants/xenobiotics.

The abstract submission deadline is July 15, 1994. Abstracts must pertain to one of the following seven program areas: recent advances in dietary assessment tools; recent advances in the evaluation of assessment methods; new approaches to assessing diets of diverse populations; advances in the evaluation and validity of specific food components; advances in methodological issues in the use of international data sets; advances in evaluating and interpreting dietary data; and the evaluation of food consumption patterns.

To receive abstract submission guidelines and/or further information, please contact the Conference on Dietary Assessment Methods, Harvard School of Public Health, 677 Huntington Ave, LL 23, Boston, MA 02115-6023; tel (617) 432-1171; fax (617) 432-1969.

Dietary Sources of Fats and Cholesterol in US Children Aged 2 through 5 Years

Frances E. Thompson, PhD, MPH, and Barbara A. Dennison, MD

ABSTRACT

Objectives. This study of lipid intakes among preschool children (1) analyzed the contributions of 38 food groups to fat, saturated fat, and cholesterol intakes; (2) estimated the effects of food substitutions on intakes; and (3) examined demographic differences in food group intake and food group sources of these lipids.

Methods. The sample consisted of 547 children, aged 2 to 5 years, from the US Department of Agriculture's 1985 and 1986 Continuing Surveys of Food Intakes by Individuals. Dietary information for 4 nonconsecutive days throughout a year was used. All foods were classified into groups and the lipids contributed from each group were computed.

Results. Over 80% of the children consumed more total fat, saturated fats, and cholesterol than is recommended. The major source of total fat and saturated fats was whole milk; the major sources of dietary cholesterol were eggs and whole milk. Children's food consumption patterns differed by region of the country and race/ethnicity, providing opportunities to refine nutrition education interventions and evaluations.

Conclusions. By substituting lower-fat foods for the major sources of saturated fats, significant reductions in preschool children's intakes of saturated fats, fat, and dietary cholesterol could be achieved. (*Am J Public Health.* 1994;84:799–806)

Introduction

Numerous private and governmental organizations have recommended that children over 2 years of age restrict dietary fats and cholesterol.[1–3] Recent recommendations issued by the National Cholesterol Education Program's Expert Panel on Blood Cholesterol Levels in Children and Adolescents call for limiting the intake of saturated fatty acids to less than 10% of total calories, total fat to no more than 30% of total calories, and dietary cholesterol to less than 300 mg/day for children aged 2 years and older.[4] The American Academy of Pediatrics Committee on Nutrition concurred with these recommendations in 1992.[5] The recommendations are consistent with the *Dietary Guidelines for Americans* issued by the US Department of Agriculture (USDA) and the US Department of Health and Human Services.[6]

The dietary patterns of specific regional populations of young children have been examined.[7–9] We use dietary data from nationwide samples of US preschool-aged children to (1) estimate for various demographic subgroups the prevalence with which dietary lipid recommendations are met; (2) describe major food group sources of dietary fat, saturated fatty acids, and cholesterol; (3) estimate the effects of food substitutions on lipid intakes; and (4) analyze differences in food group consumption by various demographic variables.

Methods

Study Design

Data from USDA's 1985 and 1986 Continuing Surveys of Food Intakes by Individuals core sample were used. The design of these surveys provided two multistage stratified area probability samples of women aged 19 through 50 years and their children aged 1 through 5 years that were representative of the 48 conterminous states. Stratification variables of the sampling design included geographical location, degree of urbanization, and socioeconomic characteristics.

Although the same area segments were sampled in both the 1985 and 1986 surveys, different housing units were selected for each survey. To ensure comparability of the two surveys, data collection and processing procedures used were identical.

Assessment of Dietary Intake

Respondents were asked to provide six nonconsecutive 1-day dietary recalls collected at approximately 2-month intervals over a 1-year period. Interviews were conducted from April 1985 through March 1986 and from April 1986 through March 1987. In eligible households, all women aged 19 through 50 years at the time of initial contact and their children aged 1 through 5 years were asked to participate. Participating households were scheduled for interviews on different days of the

At the time of the study, Frances E. Thompson was with the Division of Nutritional Sciences, Cornell University, Ithaca, NY. She is now with the Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Md. Barbara A. Dennison is with the Department of Pediatrics, College of Physicians and Surgeons of Columbia University, New York, NY, and the Research Institute and Department of Pediatrics, Mary Imogene Bassett Hospital, Cooperstown, NY.

Requests for reprints should be sent to Frances E. Thompson, PhD, MPH, Executive Plaza North, Room 313, National Cancer Institute, 9000 Rockville Pike, Bethesda, MD 20892.

This paper was accepted October 14, 1993.

Editor's Note. See related editorial by Nestle (p 713) in this issue.

week, including weekends. For subsequent contacts, interviewers attempted to reinterview within 10 days of 2 months' time from the date of the last interview, and on a different day of the week. Each participating woman provided dietary intake information for herself and her children. The first interview was conducted in person; subsequent interviews were conducted by telephone. Standard interviewing methods were used, including probing for forgotten foods,[10,11] and standard household measuring cups and spoons and a ruler were used to assist respondents in reporting portion sizes.

Study Sample

Although the design of the surveys called for collection of 6 nonconsecutive days of dietary information, fewer than half of the respondents interviewed initially provided 6 days of dietary information. Consequently, USDA constructed a sample of all respondents who had dietary data for at least 4 of the 6 days. For this sample, the first day of information was included for all respondents. For those who had 5 or 6 days of dietary information, USDA randomly selected 3 additional days to be used in the data set. Of those children with food intake data from the initial interview, 62% of the 1985 core sample and 68% of the 1986 core sample had at least 3 additional days of dietary information. The resultant data set from the combined 1985 and 1986 core samples included 649 children, each with 4 days of dietary information. The study procedures and nonresponse analysis are described in detail elsewhere.[10–12]

Because the current dietary recommendations do not pertain to children younger than 2 years, 1-year-old children were omitted from these analyses, leaving 547 children in the sample. When population weights were applied, the children were distributed similarly across 1-year age groups. About 81% were White, 10% Black, 6% Hispanic, and 3% other. Eighteen percent lived in the Northeast, 24% in the Midwest, 33% in the South, and 25% in the West. Twenty-eight percent had household incomes at or below 130% of the federal poverty level, 45% had incomes between 130% and 300% of the poverty level, and 27% had incomes above 300% of the poverty level. These characteristics were distributed similarly across sex.

Variable Definition

Individual food items were classified into 38 food groups on the basis of similar patterns of use, frequency of use, and nutrient composition. (A description of the food groups is available from Dr Thompson.) Foods eaten together and reported as one food item (e.g., a beef sandwich) were categorized into mixture food groups.

For each respondent, the total amount of each food group consumed (in grams) was determined. In addition, the amounts of each of the three lipids highlighted by the National Cholesterol Education Program's pediatric panel—dietary total fat, saturated fatty acids, and cholesterol—consumed from each food group were computed.

The contribution of each food group to the intake of each lipid was examined in two ways: as the population proportion and as the mean proportion. The population proportion was estimated as follows:

$$\text{Population Proportion} = \frac{\sum_{i=1}^{n} F_i}{\sum_{i=1}^{n} T_i},$$

where F_i = the amount of lipid contributed by a particular food group for the ith individual, T_i = the total amount of lipid from all foods from the ith individual, and n = the number of individuals in the population.

With cholesterol from eggs as an example, this measure represents the proportion of all cholesterol consumed by the population that comes from eggs. Because a true measure of usual intake is unobtainable, this is the closest approximation.[13] However, because it is measured for the whole population, there is no way to examine the variance for this proportion or to examine it in relation to other variables. Therefore, the mean proportion was estimated as follows:

$$\text{Mean Proportion} = \frac{\sum_{i=1}^{n} \frac{F_i}{T_i}}{n},$$

where the terms are defined as previously.

Again using cholesterol from eggs as an example, the mean proportion represents the average contribution of eggs to cholesterol per person. Because the proportion (F_i/T_i) is calculated for each person individually, a standard deviation can be determined and the variable can be examined in relation to other measures. The use of multiple days of data legitimizes this approach.[13]

Analysis

Weighting factors to adjust for household and individual nonresponse were incorporated in all analyses. Proportions and means were estimated with SAS software (SAS Institute Inc, Cary, NC); standard errors and statistical significance were estimated with SUDAAN software (Research Triangle Institute, Research Triangle Park, NC).

Multiple regression analysis was used to examine the independent effects of various demographic characteristics on food group consumption (in grams per day). Independent variables included region of residence, household income (percentage of poverty level, adjusted for family size and household composition), sex, age, and race/ethnicity. Weighted least squares means were estimated for various subgroups: the weighted REG procedure of SAS was used to derive estimates of regression coefficients and then the weighted means for each category of interest were computed. Regression coefficients for continuous variables and statistical significance for all variables in the model were estimated through the REGRESS procedure of SUDAAN.

Results

Macronutrient and overall energy intakes increase with age (Table 1), ranging from 1371 kcal at age 2 years to 1539 kcal at age 5 years. However, cholesterol intake, percentage of calories from total fat, and percentage of calories from saturated fatty acids are similar across age groups, averaging 232 mg/day, 35%, and 14%, respectively.

Nearly all children aged 2 through 5 years consume diets with more than the recommended amounts of total fat and saturated fatty acids (Table 2). Although 75% of children consume less than 300 mg/day of cholesterol, 81% consume diets with more than 100 mg of cholesterol per 1000 kcal, an alternative standard for lower-energy diets.[14] Black children and children from the lowest-income households have diets least consistent with dietary fat and cholesterol recommendations.

Food Group Sources of Dietary Lipids

Total fat. For the total population of children aged 2 through 5 years, the major food group source of total dietary fat is whole milk, which contributes an estimated 6.4 g/day of fat, 11.3% of the total amount of fat consumed (Table 3). Sweet

TABLE 1—Weighted Mean Daily Nutrient Intakes[a] of US Preschool-Aged Children, by Age

Nutrient	Age										
	2 y (n = 124)		3 y (n = 158)		4 y (n = 127)		5 y (n = 138)		2–5 y (n = 547)		
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	P[b]
Energy, kcal	1371	39	1446	33	1492	43	1539	44	1462	23	<.005
Protein, g	52.1	1.4	53.5	1.6	56.3	1.8	57.8	1.9	54.9	0.9	<.01
Carbohydrate, g	175.3	5.2	187.3	4.5	191.7	5.6	198.7	5.6	188.3	3.2	<.001
Total fat, g	53.7	2.2	56.2	1.6	58.3	2.1	59.4	2.2	56.9	1.2	<.05
Saturated fatty acids, g	21.5	0.9	21.9	0.6	23.1	0.9	23.4	0.9	22.5	0.5	<.05
Cholesterol, mg	228	11	223	10	238	11	239	12	232	6	NS
% calories from total fat	35.1	0.6	34.8	0.5	35.0	0.5	34.5	0.5	34.9	0.4	NS
% calories from saturated fatty acids	14.1	0.3	13.6	0.2	13.9	0.2	13.6	0.2	13.8	0.2	NS
Cholesterol, mg/1000 kcal	170	9.4	155	5.5	162	6.9	154	6.9	160	4.0	NS

Note. NS = not significant.
Source. Data are from the US Department of Agriculture, Human Nutrition Information Service.[10,11]
[a]Mean of four nonconsecutive 1-day dietary recalls for each individual, averaged over all individuals in the group, weighted to represent population.
[b]Statistical significance of slope tested by simple regression (nutrient = $a_0 + b_1$ age), taking into account weighting and design effects.

TABLE 2—Percentage Distribution of US Children Aged 2 through 5 Years According to Dietary Intake of Total Fat, Saturated Fatty Acids, and Cholesterol, by Demographic Characteristics

		% Calories from Total Fat				% Calories from Saturated Fatty Acids	Dietary Cholesterol	
	n	≤30.0	30.1–35.0	35.1–40.0	>40.0	<10.0	<300 mg/day	≤100 mg/1000 kcal
Sex								
Male	273	16.3	37.3	33.1	13.3	8.1	73.9	17.4
Female	274	14.5	40.0	31.0	14.6	3.1	78.7	20.3
Age, y								
2	124	12.1	43.7	28.0	16.1	10.3	81.0	11.6
3	158	18.4	30.5	37.4	13.7	7.5	79.8	20.5
4	127	14.2	42.3	29.4	14.1	2.2	71.6	20.3
5	138	16.5	38.8	32.7	12.0	2.3	72.6	22.8
Race/ethnicity[a]								
White	454	16.6	37.1	31.4	14.9	6.0	77.5	22.1
Black	46	3.0	45.1	40.3	11.6	1.5	80.8	2.1
Hispanic	32	18.4	44.2	26.2	11.3	7.4	53.4	3.8
Region								
Northeast	91	21.0	46.8	23.9	8.3	7.4	71.3	23.6
Midwest	163	13.9	37.6	37.8	10.7	3.3	80.9	22.7
South	164	12.8	36.8	33.3	17.2	6.1	76.1	13.9
West	129	16.2	36.1	30.7	17.0	5.7	75.9	18.4
Household income								
≤130% of poverty level	166	9.6	31.7	39.4	19.3	2.8	70.5	8.9
131–300% of poverty level	252	18.3	45.3	26.3	10.1	7.9	77.9	22.6
>300% of poverty level	129	16.7	34.9	33.7	14.7	4.5	79.8	23.2
Total	547	15.4	38.6	32.0	14.0	5.6	76.3	18.9

Source. Data are from the US Department of Agriculture, Human Nutrition Information Service.[10,11]
[a]Data for "other" racial/ethnic groups (n = 15) not given.

grain products (e.g., cookies, cakes, pies, doughnuts); franks, sausages, lunch meats, and bacon; and mostly grain mixtures (e.g., macaroni and cheese, pizza, Hispanic mixtures) contribute an estimated 4.8, 4.6, and 4.1 g/day of fat, respectively. Sixteen food groups account for 85.6% of the fat consumed by the population.

Saturated fats. The major source of saturated fatty acids is the whole milk food group, which contributes an estimated 4.0 g/day of saturated fatty acids, 17.8% of the total amount of saturated fatty acids consumed by children aged 2 through 5 years (Table 3). Natural and

TABLE 3—Major Food Group Sources of Dietary Total Fat, Saturated Fatty Acids, and Cholesterol for US Children Aged 2 through 5 Years

Food Group	Total Fat			Saturated Fatty Acids			Cholesterol		
	Rank	g/day	% of Total	Rank	g/day	% of Total	Rank	mg/day	% of Total
Whole milk	1	6.4	11.3	1	4.0	17.8	2	26	11.2
Sweet grain products	2	4.8	8.5	6	1.5	6.5	3	18	7.9
Franks, sausages, lunch meats, bacon	3	4.6	8.1	3	1.7	7.6	7	10	4.5
Mostly grain mixtures	4	4.1	7.3	4	1.7	7.4	5	15	6.4
Nut butters, nuts, seeds	5	3.3	5.8	14	0.5	2.4	...[a]	...[a]	...[a]
Natural and processed cheese	6	2.9	5.2	2	1.8	8.2	10	9	3.8
Beef (reported separately)	7	2.9	5.1	8	1.2	5.3	4	16	6.7
Butter and margarine	8	2.8	4.9	9	1.0	4.2	...[a]	...[a]	...[a]
Salty snacks	9	2.6	4.6	11	0.8	3.4	...[a]	...[a]	...[a]
Low-fat milks	10	2.6	4.6	5	1.6	7.3	8	10	4.5
White potatoes	11	2.2	3.8	12	0.7	3.3	...[a]	...[a]	...[a]
Meat mixtures	12	2.1	3.6	10	0.8	3.5	9	9	4.0
Dairy and frozen desserts	13	2.0	3.5	7	1.2	5.6	11	8	3.3
Eggs	14	2.0	3.5	13	0.6	2.7	1	70	30.3
Poultry	15	1.8	3.2	15	0.5	2.2	6	13	5.6
Bread	16	1.5	2.6	...[a]	...[a]	...[a]	...[a]	...[a]	...[a]
Sugar and candy	...[a]	...[a]	...[a]	16	0.4	2.0	...[a]	...[a]	...[a]
Cured, organ, miscellaneous meats	...[a]	...[a]	...[a]	...[a]	...[a]	...[a]	...[a]	...[a]	...[a]
Fish	...[a]	...[a]	...[a]	...[a]	...[a]	...[a]	12	5	2.2
Total		48.6	85.6		20.0	89.4		209	90.4

Source. Data are from the US Department of Agriculture, Human Nutrition Information Service.[10,11]
[a]Provides less than 2% of total nutrient.

processed cheese; franks, sausages, lunch meats, and bacon; and mostly grain mixtures contribute 1.8, 1.7, and 1.7 g/day, respectively, of saturated fatty acids. Sixteen food groups account for 89.4% of the saturated fatty acids consumed by the population.

Dietary cholesterol. Almost a third (30.3%; 70 mg/day) of the total amount of cholesterol consumed by children aged 2 through 5 years comes from the egg food group (Table 3). The food groups whole milk, sweet grain products, and beef (reported separately) contribute an additional 26, 18, and 16 mg/day, respectively, of dietary cholesterol. Twelve food groups account for 84.4% of the dietary cholesterol consumed by the population.

Effects of Food Substitutions on Lipid Intakes

By targeting children's major dietary sources of saturated fatty acids, one can examine the effects of potential changes. Whole milk, natural and processed cheese, and low-fat milks (e.g., skim, 1%, or 2% milk) contribute 33% of the saturated fatty acids in these children's diets (Table 3). Substituting 1% milk for all the whole milk consumed (cup for cup), part-skim-milk cheese for all the whole-milk natural cheese consumed (ounce for ounce), and skim milk for all of the low-fat milks consumed (cup for cup) would decrease the population's intake of saturated fatty acids (from 22.5 to 16.7 g/day), total fat (from 56.9 to 47.3 g/day), and cholesterol (from 209 to 174 mg/day). The 5.8-g/day reduction in saturated fatty acid intake is more than 90% of the 6.3-g/day reduction needed to achieve an average saturated fatty acid intake of 16.2 g/day (10% of total calories). These substitutions also result in a 9.6-g/day reduction in total fat, more than the 8.2-g/day reduction needed to achieve a total fat intake of 48.7 g/day (30% of total calories). These substitutions of lower-fat foods for higher-fat foods would reduce total energy intake from 1462 to 1363 calories. To compensate for this deficit, the diets can be made isocaloric by increasing the intake of low-fat complex carbohydrate food groups (grains, vegetables, and fruits), consistent with federal dietary guidelines.[3,4,6]

The food groups (1) franks, sausages, lunch meats, and bacon and (2) mostly grain mixtures are also among the top five food group sources of dietary saturated fatty acids. Reduced-fat turkey hot dogs or lunch meats or lean cuts of beef, poultry, or fish could be substituted for regular hot dogs or lunch meats. Reduced-fat versions of many grain mixtures are also available. In home-prepared versions, the total or saturated fat content could be reduced by using less fats and oils, substituting mono- or polyunsaturated fat for saturated fat, and substituting lower-fat cheese for regular cheese.

The population-level data presented guide the design and evaluation of nutrition education efforts for populations. Recommendations for particular individuals must be tailored to individual tastes and constraints. Not all individuals will find a given substitution acceptable and not all will use it all the time. Moderate changes in the major sources of saturated fatty acids for children will result in significant favorable changes in dietary intake for most children. When the dietary fat intake of young children is reduced, extremes need to be avoided, as very-low-fat diets with inadequate calories and nutrients may result.[15]

Demographic Differences in Food Group Consumption

The effects of various demographic variables on the consumption of selected food groups, with the effects of the other variables in the model controlled, are

TABLE 4—Weighted Least Squares Means and Regression Coefficients of Demographic Variables and Food Group Consumption (g/day) for US Children Aged 2 through 5 Years[a]

Food Group	Least Squares Means: Region					Least Squares Means: Race/Ethnicity				Least Squares Means: Sex		Beta Coefficients	
	Northeast (n = 88)	Midwest (n = 161)	South (n = 162)	West (n = 121)	Significance[b]	White (n = 454)	Black (n = 46)	Hispanic (n = 32)	Significance[b]	Male (n = 266)	Female (n = 266)	Age	Household Income
Whole milk	234.0	157.8	223.1	162.7	(1, 2)* (2, 3)*	189.4	189.9	273.4	. . .	185.3	203.9	−5.7207	−0.1280*
Sweet grain products	36.0	34.7	34.5	36.2	. . .	36.7	30.1	24.3	(1, 3)**	39.1	31.4*	2.9876*	0.0105
Franks, sausages, lunch meats, bacon	12.2	22.2	18.3	18.2	(1, 2)*** (1, 3)* (1, 4)*	16.9	29.2	16.0	(1, 2)** (2, 3)*	20.5	15.7**	0.6637	−0.0085
Mostly grain mixtures	68.1	59.2	46.3	53.1	. . .	57.4	32.9	59.8	(1, 2)*	57.2	53.0	1.5856	−0.0378
Nut butters, nuts, seeds	11.4	7.9	5.7	9.1	. . .	7.2	13.7	11.5	. . .	8.6	7.6	1.1515	0.0140
Natural and processed cheese	12.0	9.8	9.4	10.0	. . .	10.8	8.3	4.8	(1, 3)***	9.8	10.5	0.0663	0.0066
Beef (reported separately)	14.8	18.0	17.6	16.8	. . .	16.1	20.3	23.3	. . .	18.9	15.1	0.9608	−0.0039
Butter and margarine	3.8	4.4	3.1	4.2	(2, 3)* (3, 4)*	4.0	2.5	2.8	(1, 2)**	3.6	3.9	0.3334*	0.0015
Salty snacks	10.8	8.8	10.4	11.7	. . .	11.4	5.6	5.4	(1, 2)*** (1, 3)**	10.5	10.3	0.4366	0.0047
Low-fat milks	125.0	224.3	91.3	221.3	(1, 2)** (1, 4)* (2, 3)*** (3, 4)**	179.5	53.2	94.1	(1, 2)*** (1, 3)*	164.9	157.2	2.7102	0.0786
White potatoes	23.6	31.3	37.5	28.2	(1, 3)*	32.2	24.8	28.5	. . .	30.4	32.0	−0.0974	0.0036
Meat mixtures	18.6	30.3	25.2	28.6	(1, 2)** (1, 4)*	27.2	18.5	23.3	. . .	26.5	25.6	0.5919	−0.0233
Dairy and frozen desserts	33.3	40.8	37.1	37.4	. . .	39.8	20.8	32.4	(1, 2)**	40.2	34.6	7.3780***	0.0214
Eggs	13.8	11.7	13.7	14.8	. . .	12.9	13.3	21.9	. . .	13.9	13.1	−0.0960	−0.0158*
Poultry	19.0	13.0	16.8	11.5	. . .	13.7	21.8	21.6	. . .	16.6	13.4	2.4534*	0.0093
Bread	45.5	36.7	36.8	38.7	. . .	39.7	36.9	30.1	(1, 3)*	41.5	36.3*	2.9236**	−0.0131
Sugar and candy	15.7	18.7	16.6	20.2	. . .	19.5	10.1	7.7	(1, 2)*** (1, 3)***	18.8	16.8	1.2302	0.0156
Cured, organ, miscellaneous meats	4.3	4.1	4.1	2.9	. . .	3.5	7.6	2.1	. . .	2.5	5.2**	0.0178	−0.0071*
Fish	4.6	4.6	5.1	7.6	. . .	4.8	11.9	3.7	(1, 2)* (2, 3)**	5.8	5.2	−0.6963	0.0093*
Total milk	359.0	382.1	314.4	384.0	(2, 3)** (3, 4)**	368.8	243.0	367.5	(1, 2)*** (2, 3)**	350.2	361.2	−3.0105	−0.0494
% of total milk as whole milk	61.9	44.1	69.5	46.5	(1, 2)* (2, 3)*** (3, 4)*	51.7	83.0	75.4	(1, 2)*** (1, 3)**	53.8	58.9	−0.1926	−0.0414**

Source. Data are from the US Department of Agriculture, Human Nutrition Information Service.[10,11]

[a]Those of "other" race/ethnicity (n = 15) excluded.

[b]Significance (two-tailed) of all pairwise contrasts. Numbers in parentheses indicate which subgroups (designated by order from left to right) are significantly different.

*$P < .05$; **$P < .01$; ***$P < .001$.

TABLE 5—Five Major Food Group Sources of Total Fat, Saturated Fatty Acids, and Dietary Cholesterol for US Children Aged 2 through 5 Years, by Region[a]

Northeast (n = 88)		Midwest (n = 161)		South (n = 162)		West (n = 121)	
Food Group	%[b]	Food Group	%[b]	Food Group	%[b]	Food Group	%[b]
			Total fat				
Whole milk	13.7	Franks, etc.	9.6	Whole milk	13.0	Whole milk	8.8
Sweet grains	8.8	Whole milk	8.9	Sweet grains	8.5	Sweet grains	8.6
Grain mixtures	8.3	Sweet grains	8.5	Franks, etc.	8.4	Franks, etc.	8.0
Cheese	7.2	Grain mixtures	8.1	Grain mixtures	6.4	Low-fat milks	7.0
Nut butters, etc.	6.8	Low-fat milks	6.9	Beef	5.5	Nut butters, etc.	6.5
			Saturated fat				
Whole milk	20.3	Whole milk	13.7	Whole milk	19.9	Whole milk	13.8
Cheese	10.3	Low-fat milks	10.6	Cheese	8.1	Low-fat milks	11.3
Grain mixtures	8.7	Franks, etc.	9.1	Franks	8.0	Franks, etc.	7.8
Sweet grains	6.7	Grain mixtures	8.2	Sweet grains	6.7	Cheese	7.7
Low-fat milks	6.0	Cheese	7.2	Grain mixtures	6.4	Sweet grains	6.9
			Cholesterol				
Eggs	20.0	Eggs	20.7	Eggs	25.2	Eggs	23.6
Whole milk	14.9	Whole milk	9.3	Whole milk	13.8	Sweet grains	11.0
Grain mixtures	9.5	Low-fat milks	8.9	Sweet grains	8.3	Whole milk	9.3
Sweet grains	8.2	Sweet grains	8.8	Beef	7.4	Low-fat milks	7.7
Poultry	6.9	Beef	8.4	Poultry	6.2	Beef	6.7

Source. Data are from the US Department of Agriculture, Human Nutrition Information Service.[10,11]
[a]Those of "other" race/ethnicity (n = 15) excluded.
[b]Percentage of total nutrient intake contributed by the food group, adjusted for sex, age, race/ethnicity, and household income.

TABLE 6—Five Major Food Group Sources of Total Fat, Saturated Fatty Acids, and Dietary Cholesterol for US Children Aged 2 through 5 Years, by Race/Ethnicity[a]

White (n = 454)		Blacks (n = 46)		Hispanics (n = 32)	
Food Group	%[b]	Food Group	%[b]	Food Group	%[b]
		Total fat			
Whole milk	10.6	Franks, etc.	14.5	Whole milk	18.1
Sweet grains	8.9	Whole milk	11.4	Grain mixtures	7.3
Grain mxtures	7.3	Sweet grains	7.3	Eggs	6.9
Franks, etc.	7.3	Beef	7.2	Franks	6.8
Nut butters, etc.	5.9	Nut butters, etc.	5.3	Beef	6.5
		Saturated fat			
Whole milk	16.1	Whole milk	18.5	Whole milk	26.9
Low-fat milks	8.7	Franks	14.4	Beef	6.9
Cheese	8.7	Beef	7.9	Grain mixtures	6.7
Grain mixtures	7.5	Cheese	6.7	Franks, etc.	6.1
Sweet grains	6.9	Sweet grains	6.3	Dairy and frozen desserts	5.2
		Cholesterol			
Eggs	22.0	Eggs	24.3	Eggs	29.9
Whole milk	11.6	Whole milk	11.2	Whole milk	15.4
Sweet grains	9.5	Beef	8.6	Beef	7.6
Beef	7.1	Poultry	7.7	Grain mixtures	7.1
Grain mixtures	7.0	Sweet grains	7.2	Sweet grains	6.6

Source. Data are from the US Department of Agriculture, Human Nutrition Information Service.[10,11]
[a]Those of "other" race/ethnicity (n = 15) excluded.
[b]Percentage of total nutrient intake contributed by the food group, adjusted for sex, age, region, and household income.

shown in Table 4. Overall, the number of significant effects is small relative to the number of tests performed. Race/ethnicity and region are associated with significant effects in more food groups than are the other variables examined.

For example, significantly less total milk (whole milk and low-fat milk groups combined) is consumed by Black children than by White and Hispanic children and in the South than in the Midwest or West. Of the total milk consumed, children in the South consume a greater proportion as whole milk (84%) than do children in the Northeast (76%), the Midwest (58%), or the West (61%). This proportion is higher in Black (82%) and Hispanic (75%) than in White children (51%). Compared with White and Hispanic children, Blacks consume significantly more franks, sausages, lunch meats, and bacon. The proportion of total milk consumed as whole milk is significantly higher for lower-income households than for higher-income households.

Demographic Differences in Food Group Sources of Lipids

For children in all racial/ethnic groups and in all regions, whole milk is the major food group source of saturated fatty

acids; eggs and whole milk are major food group sources of dietary cholesterol; and whole milk, franks and related meats, or both are the major food group sources of total fat. Differences in overall food group consumption affect which food groups are other major sources of total fat, saturated fatty acids, and dietary cholesterol for children living in different regions (Table 5) and for different racial/ethnic groups (Table 6). For example, whole milk is the major food group source of total fat for Hispanics and Whites, whereas franks and related meats constitute the major food group source for Blacks.

Discussion

Several limitations in the data used for these analyses should be noted. First, mothers may have been unaware of what foods were consumed when their children were cared for by others. Recent studies indicate that the consensus recall method, in which the child and both parents participate, may provide better intake estimates than a report from one parent alone.[16] Second, the nutrient composition database used in the 1985 and 1986 USDA surveys lacked specificity with regard to foods consumed by Hispanics.[17] Nutrient data were based on commercially prepared, Americanized versions of Mexican foods and thus did not reflect traditional preparation methods or differences in preparation methods among Hispanic subgroups.

The small number of children in certain subgroups (e.g., Blacks [n = 46] and Hispanics [n = 32]) precluded testing for important statistical interactions involving race/ethnicity. For example, Hispanics living in the South and West (most of whom come from Mexico) may have different diets than Hispanics living in the Northeast (most of whom come from Puerto Rico). In addition, many statistical tests were conducted for the demographic comparisons, increasing the chances of detecting random rather than real effects. Thus, these results are more appropriately viewed as descriptive and suggestive of the need for corroboration from other work. However, results from other studies are consistent with our work. For example, our finding that Hispanics drink substantially more whole milk than low-fat milk is consistent with findings from other studies.[8,9,18] Similarly, our finding that more Black than White children and more lower-income households than higher-income households had diets that failed to meet dietary lipid recommendations is consistent with analyses of the second National Health and Nutrition Examination Survey.[19,20]

Finally, substantial secular changes in food consumption may have occurred since 1985 and 1986. More lower-fat food products are now offered in the marketplace. Consumption of whole milk decreased and that of low-fat milk increased between 1985 and 1990.[21] The effects of these trends on the dietary patterns of children should be examined as newer data become available.

Although numerous programs and initiatives provide dietary guidance for school-aged children,[22–24] comprehensive dietary guidance provided for preschool-aged children is limited. Recommendations are available in the National Cholesterol Education Program report.[4] USDA and the Department of Health and Human Services provide qualitative dietary guidance to Child Nutrition Program cooperators, including day-care providers.[25] The Supplemental Food Program for Women, Infants, and Children (WIC), which provides food vouchers for eligible low-income infants and children to age 5, offers opportunities to support consumption of low-fat products, both in choice of acceptable foods for vouchers and in nutrition education for clients.

Our results showing the importance of dairy products, particularly whole milk, to lipid consumption highlight the need to promote low-fat dairy products comprehensively throughout all food and nutrition efforts aimed at children, from preschool through adolescence. Analyses of differences in food consumption patterns among different demographic subgroups indicate that certain relationships are strong for all examined subgroups. However, differences in food group consumption do occur, and these differences should be incorporated into the design of nutrition education programs. Baseline assessment of dietary sources of lipids at the regional level may lead to more effective interventions and valid program evaluations. □

Acknowledgments

This study was carried out through the support of the Human Nutrition Information Service, US Department of Agriculture, Cooperative Agreement 58-3198-6-34 and the New York State Department of Health, Bureau of Nutrition, Contract C001038. The research was conducted using the Cornell National Supercomputer Facility, a resource of the Cornell Theory Center, which is funded in part by the National Science Foundation, New York State, the IBM Corporation, and members of the Cornell Theory Center's Corporate Research Institute.

The authors gratefully acknowledge research assistance from Holly Gump; data and computer assistance from Sharon Bushart, Edward A. Frongillo, Jr, Marc Melitz, Lisa Kahle, and Joseph Goldman; and helpful comments from Cutberto Garza, Barry I. Graubard, Susan M. Krebs-Smith, Rachel Ballard-Barbash, and Anne Shaw.

References

1. *NIH Consensus Development Statement on Lowering Blood Cholesterol to Prevent Heart Disease.* Vol. 5, no. 7. Washington, DC: National Institutes of Health; 1984.
2. American Health Foundation, EL Wynder, ed. Coronary artery disease prevention: cholesterol, a pediatric perspective. *Prev Med.* 1989;18:323–409.
3. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50212.
4. National Cholesterol Education Program. *Report of the Expert Panel on Blood Cholesterol Levels in Children and Adolescents.* Bethesda, MD: US Dept of Health and Human Services; September 1991. DHHS publication NIH 91-2732.
5. American Academy of Pediatrics Committee on Nutrition. Statement on cholesterol. *Pediatrics.* 1992;90:469–473.
6. *Nutrition and Your Health: Dietary Guidelines for Americans.* 3rd ed. Washington, DC: US Dept of Agriculture and US Dept of Health and Human Services; 1990.
7. Nicklas TA, Farris RP, Hyg MS, et al. Cardiovascular risk factors from birth to 7 years of age: The Bogalusa Heart Study. Dietary intakes. *Pediatrics.* 1987;80(suppl): 797–816.
8. Murphy SP, Castillo RO, Martorell R, Mendoza FS. An evaluation of food group intakes by Mexican-American children. *J Am Diet Assoc.* 1990;90:388–393.
9. Basch CE, Shea S, Zybert P. Food sources, dietary behavior, and the saturated fat intake of Latino children. *Am J Public Health.* 1992;82:810–815.
10. *Nationwide Food Consumption Survey Continuing Survey of Food Intakes by Individuals, Women 19–50 Years and Their Children 1–5 Years, 4 Days, 1985.* Washington, DC: US Dept of Agriculture, Human Nutrition Information Service; 1987. NFCS, CSFII Report no. 85-4.
11. *Nationwide Food Consumption Survey Continuing Survey of Food Intakes by Individuals, Women 19–50 Years and Their Children 1–5 Years, 4 Days, 1986.* Washington, DC: US Dept of Agriculture, Human Nutrition Information Service; 1988. NFCS, CSFII Report no. 86-3.
12. Tuszynski C, Roidt L. *Analysis of Nonresponse to the 1986 Continuing Survey of Food Intakes by Individuals.* Washington, DC: US Dept of Agriculture, Human Nutrition Information Service; 1988. Administrative Report no. 386.
13. Krebs-Smith SM, Kott PS, Guenther PM. Mean proportion and population proportion. Two answers to the same question? *J Am Diet Assoc.* 1989;89:671–676.

14. Weidman W, Kwiterovich P Jr, Jesse MJ, Nugent E. Diet in the healthy child: Task Force Committee of the Nutrition Committee and the Cardiovascular Disease in the Young Council of the American Heart Association. *Circulation.* 1983;74:1411A–1414A.
15. Sigman-Grant M, Zimmerman S, Kris-Etherton PM. Dietary approaches for reducing fat intake of preschool-age children. *Pediatrics.* 1993;91:955–960.
16. Eck LH, Klesges RC, Hanson CL. Recall of a child's intake from one meal: are parents accurate? *J Am Diet Assoc.* 1989;89:784–789.
17. Loria CM, McDowell MA, Johnson CL, Woteki CE. Nutrient data for Mexican American foods: are current data adequate? *J Am Diet Assoc.* 1991;91:919–922.
18. Frank GC, Zive M, Nelson J, Broyles SL, Nader PR. Fat and cholesterol avoidance among Mexican-American and Anglo preschool children and parents. *J Am Diet Assoc.* 1991;91:954–958, 961.
19. Kimm SYS, Gergen PJ, Malloy M, Dressler C, Carroll M. Dietary patterns of U.S. children: implications for disease prevention. *Prev Med.* 1990;19:432–442.
20. Martinez GA, Ryan AS, Rosche AF. What are children in the U.S. eating? In: Roche AF, ed. *Prevention of Adult Atherosclerosis during Childhood.* Report of the Ninety-Fifth Ross Conference on Pediatric Research. Columbus, Ohio: Ross Laboratories; 1988:53–60.
21. Putnam JJ, Allshouse JE. *Food Consumption, Prices, and Expenditures, 1970–1990.* Washington, DC: US Dept of Agriculture, Economic Research Service; 1992. Statistical Bulletin no. 840.
22. *The Healthy E.D.G.E. in Schools: Eating, the Dietary Guidelines and Education.* Alexandria, Va: School Food Service Foundation of the American School Food Service Association; 1991.
23. Stone EJ. School-based health research funded by the National Heart, Lung, and Blood Institute. *J School Health.* 1985;55:168–174.
24. Stone EJ, Perry CL, Luepker RV. Synthesis of cardiovascular behavioral research for youth health promotion. *Health Ed Q.* 1989;16:155–169.
25. *Building for the Future: Nutrition Guidance for the Child Nutrition Programs.* Alexandria, Va: US Dept of Agriculture and US Dept of Health and Human Services, Nutrition and Technical Services Division; 1992. FNS-279.

Overweight Schoolchildren in New York State: Prevalence and Characteristics

Wendy S. Wolfe, PhD, Cathy C. Campbell, PhD, Edward A. Frongillo, Jr, PhD, Jere D. Haas, PhD, and Thomas A. Melnik, DrPH

ABSTRACT

Objectives. Childhood overweight is an increasing public health concern. This study was undertaken to determine the prevalence of overweight in elementary school children in New York State and to identify characteristics associated with child fatness.

Methods. Weight, height, triceps skinfold, midarm circumference, and a 24-hour dietary recall were taken on 1797 second- and fifth-grade students from 51 randomly selected schools in New York State outside of New York City. Parents completed a brief questionnaire.

Results. In comparison with 1974 and 1980 national reference data, up to twice the expected percentages of children had values above the 85th, 90th, and 95th percentiles for body mass index, triceps skinfold, and arm fat area. Regression analyses suggested that children who tended to be fatter were members of low socioeconomic status, two-parent (but not single-parent) households; those with few or no siblings; those who ate school lunch; and those who skipped breakfast.

Conclusions. The findings suggest that overweight is a problem among many elementary school children in New York State and that sociodemographic characteristics may be useful for targeting preventive efforts. (*Am J Public Health.* 1994;84: 807–813)

Introduction

Childhood overweight is an increasing public health concern, affecting as many as 20% to 25% of children in the United States.[1–3] According to one analysis, the prevalence of overweight in 6- to 11-year-old children increased by 54% from the early 1960s to the late 1970s.[2]

Overweight children have an increased risk of adult obesity that increases with the age of the child and the severity of the overweight.[4] In longitudinal studies, 27% of overweight 1- to 5-year-olds, 41% to 43% of overweight 3- to 9-year-olds, and 80% to 86% of overweight 10- to 13-year-olds have remained overweight as adults.[5–7] Overweight children also have increased risk factors for heart disease and diabetes and increased emotional stress, orthopedic disorders, and respiratory disease.[8]

The present study was undertaken to determine the prevalence of overweight and underweight in elementary school children in New York State outside of New York City and to identify characteristics associated with child fatness.

Methods

Sample Selection and Representativeness

A two-stage sampling technique was used. First, schools were randomly selected within 21 strata: 7 geographical regions in New York State outside of New York City and 3 socioeconomic status (SES) levels within each of these regions (based on 1980 census data). Of the 110 schools contacted by letter, 51 (46%) volunteered to participate. Parent consent forms with brief questionnaires were given to second and fifth graders; 1797 (51%) agreed to participate and were surveyed. Ages ranged from 6 to 12 years; the mean age of second graders was 7.9 (SD = 0.5), and the mean age of fifth graders was 10.9 (SD = 0.5).

Comparison of the students surveyed with data on all children in New York State outside of New York City suggested that the sample was representative.[9] As in the population, the sample was predominantly White, 18% lived with a single parent, and just over half had mothers who were employed.

Nonparticipation of schools did not appear to relate to the prevalence of overweight. Reasons included small school size, factors related to the dental component of the study,[10] and a misperception that the survey was intended only for low-income children.

Within the 51 schools, only two samples differed significantly from their school population in the proportion of non-White students, and two others differed in the proportion of students eligible for free or reduced-price school lunches. Because this number of differences would be expected as a result of sampling error, no schools were omitted from the analysis.[11]

Wendy S. Wolfe, Edward A. Frongillo, Jr, and Jere D. Haas are and Cathy C. Campbell was with the Division of Nutritional Sciences, Cornell University, Ithaca, NY. Thomas A. Melnik is with the Division of Nutrition, New York State Department of Health, and the School of Public Health, State University of New York at Albany.

Requests for reprints should be sent to Wendy S. Wolfe, PhD, Division of Nutritional Sciences, MVR Hall, Cornell University, Ithaca, NY 14853-4401.

This paper was accepted September 1, 1993.

Editor's Note. See related editorial by Nestle (p 713) in this issue.

TABLE 1—Prevalence Estimates (%) of Overweight in the Study Sample and in the Northeast NHANES II Subsample

	New York State Outside New York City, 1987/88					Northeast United States 1976–1980 (NHANES II)	
	NHANES I			NHANES II			
Percentile	BMI	TSF	AFA	BMI	TSF	BMI	TSF
>85th	28	23	NA	23	19	25	24
>90th	NA	NA	19	17	14	20	15
>95th	13	13	12	10	8	8	8

Note. Variation depended on the reference data used for comparison (National Health and Nutrition Examination Survey [NHANES] I [1971–1974][20] or NHANES II [1976–1980][15]), the cutoff level (percentile) used to define overweight, and the indicator used to measure child fatness (body mass index [BMI], triceps skinfold [TSF], or arm fat area [AFA]).

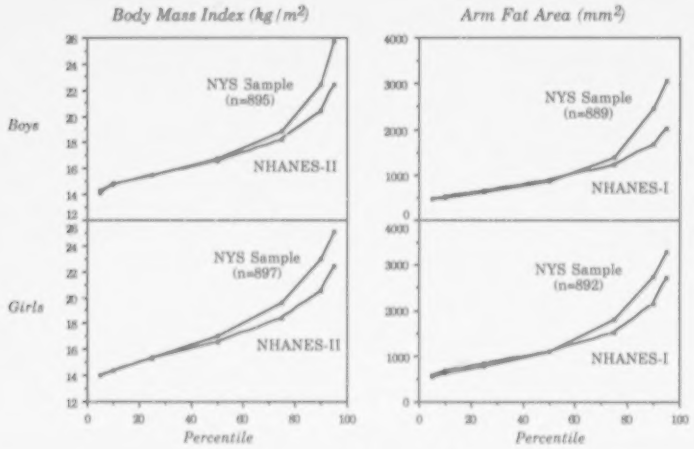


FIGURE 1—Sample arm fat area and body mass index cumulative distributions compared with the first and second National Health and Nutrition Examination Survey (NHANES I[14] and NHANES II[15]) national reference distributions, by sex.

Data Collection and Measurements

The best way to measure child fatness in the field is unclear. Various authors have recommended body mass index (weight/height^2),[12,13] triceps skinfold,[13] or arm fat area[14] (calculated from triceps skinfold and mid-upper-arm circumference: arm fat area [mm^2] = arm area [A] − arm muscle area [M], where $A = \pi/4 \times [C/\pi]^2$ and $M = [C - \pi T]^2/4\pi$ [C = arm circumference, T = triceps skinfold]) as the best indicator of body fat in school-aged children.

In this study, body mass index, triceps skinfold, and arm fat area were all used to measure child fatness. One of two trained examiners used standard procedures to measure weight, height, mid-upper-arm circumference, and triceps skinfold on each child. The principal examiner (first author) measured 1615 children between November 1987 and March 1988 and also administered a nonquantitative 24-hour dietary recall. The other examiner measured 182 children in the spring of 1987.

Measurement errors[16] for the four anthropometric indices were acceptable. Intraobserver (test–retest) reliabilities for each examiner ranged from 97% to 100%. Interobserver reliabilities, based on independent samples, were 93% to 100% between the principal examiner and an anthropometric expert and 96% to 100% between the principal and secondary examiners.

The best way to define overweight or underweight in children is also unclear. Therefore, various percentile cutoffs, as well as reference data from both the first and second National Health and Nutrition Examination Surveys (NHANES I and NHANES II), were used for comparison purposes in deriving prevalence estimates. In addition, because obesity rates are known to be higher in the Northeast than in other regions of the United States,[2,17] prevalence estimates for 6- to 12-year-olds in the Northeast subsample of NHANES II, with the total NHANES II distribution as the reference, also were calculated. Logistic regression was used to adjust the seasonality of this subsample to match that of the New York sample (November through March), since prevalence rates also vary with season.[2,17]

Five diet quality indices were derived from the recall: food diversity (number of different food items eaten), skipping of breakfast, consumption of no vegetable other than potatoes or tomato sauce, number of snack foods eaten, and a food group pattern score ("poor" defined as missing one or more of the four food groups or consuming two or more groups only one time each). Children who skipped breakfast were asked whether this was typical; most said it was.

On the basis of the parent questionnaire, children were classified as being of low SES if they met at least one of the following criteria: no parent working, eligibility for free or reduced-price school lunches, or receipt of social or food assistance. Children were classified as medium/high SES only if they met none of these criteria and had values for all three variables. Twenty-six percent of the children were considered low SES. In schools involved in the National School Lunch Program, children were classified as participants if their parent responded positively to the question "Does your child eat school lunch?"

Statistical Analysis

Regression models were developed to identify factors characterizing fatter from leaner children after adjustment for possible confounding and interactive effects. Factors examined were height, grade, age within grade, sex, race (White or African American), SES, family structure, school lunch participation, number of siblings, maternal employment, and the five dietary indices. ("Whites" included the small number of Hispanics and "others";

the body fatness of these groups did not differ significantly from that of Whites.) Because all variables were examined simultaneously for their effects after adjustment for the others, children with missing values for any of the variables, including those in nine schools that did not participate in the National School Lunch Program and those in six schools for which dietary data were not available, were omitted from the regression analyses.

Regression models were developed for each of six different anthropometric dependent variables. Two ordinary least squares models were developed for body mass index and arm fat area treated as continuous variables. Because their distributions were positively skewed and their variances increased with age and height, both were transformed into natural logarithms to make the regression residuals more symmetrically distributed.

To test for the risk of overweight or underweight, four additional models were examined with multiple logistic regression; body mass index and arm fat area were dichotomized into (a) overweight (>90th percentile) vs normal weight (≤90th and >10th percentile) and (b) underweight (≤10th percentile) vs normal weight. Body mass index values were compared with age- and sex-specific reference percentiles from the NHANES II national probability sample survey,[15] while arm fat area values were compared with reference percentiles developed by Frisancho from NHANES I.[14] Although derived from different surveys, these were the only published national reference percentiles available for body mass index and arm fat area when the regression analyses were done. For consistency in the analyses, the 90th percentile was selected as the cutoff for both indices (the 85th percentile was not available for arm fat area).

For the least squares analyses, a stepwise multiple regression procedure was run in which all main effects were retained and each conceptually plausible two-way interaction effect was tested; the least significant interaction was removed at each step, and those with a P value of .10 or less were retained.[18,19] This procedure was repeated retaining all significant two-way interactions and their inclusive main effects and testing the remaining main effects (retaining those with a P value of less than .05). The resulting models were then tested with logistic regression; variables were removed by means of a similar stepwise procedure.

TABLE 2—Sample Distributions of Physical Indices, Compared with Age- and Sex-Specific National Reference Percentiles

Indices and Percentiles	Boys (n = 900), %	Girls (n = 897), %	Second Graders (n = 960), %	Fifth Graders (n = 837), %	Total (n = 1797), %
Body mass index[a]					
≤5th	2.3	5.0	4.0	3.2	3.7
≤10th	7.0	8.2	6.9	8.3	7.7
>85th	20.9	24.5	23.0	22.4	22.6
>90th	15.6	18.3	18.3	15.5	16.9
>95th	8.7	10.9	8.8	8.3	9.8
Triceps skinfold[a]					
≤5th	4.5	5.5	4.9	5.2	5.0
≤10th	10.7	12.5	11.3	12.0	11.6
>85th	15.9	21.8	19.0	18.8	18.9
>90th	12.0	16.3	13.4	15.1	14.2
>95th	7.4	9.6	10.7	8.5	8.5
Arm fat area[b]					
≤5th	4.3	4.8	3.3	5.9	4.5
≤10th	8.4	11.4	9.7	10.1	9.9
>90th	18.3	20.1	17.9	20.8	19.3
>95th	13.2	11.7	13.7	11.1	12.5
Midarm circumference[a]					
≤5th	4.3	4.1	4.1	4.4	4.2
>95th	8.9	10.0	10.5	8.2	9.4
Height for age[c]					
≤5th	4.0	3.5	3.2	4.5	3.8
>95th	7.2	6.0	6.0	5.8	6.6

[a]Compared with the second National Health and Nutrition Examination Survey (NHANES II).[15]
[b]Compared with NHANES I.[14]
[c]Compared with National Center for Health Statistics percentiles.[21]

Results

Prevalence Estimates for Overweight and Underweight

The estimated prevalence of overweight in the sample (Table 1) varied with (a) the reference data used for comparison,[15,20] (b) the cutoff level used to define overweight, and (c) the indicator used to measure child fatness. Regardless, the prevalence was higher (by as much as twice) than that expected based on the US reference populations. Prevalence estimates differed little from those for the Northeast subsample of NHANES II, however.

The sample anthropometric distributions are shown in Figure 1 and Table 2.[14,15,21] A higher than expected percentage of children had values in the upper percentiles for body mass index, triceps skinfold, and arm fat area, while the percentage with values in the lower extremes (underweight) was similar to or less than expected. Weight status did not vary by grade or sex; within these categories, body mass index measurements were correlated significantly with both arm fat area (coefficients of .92 to .94) and triceps skinfold (.87 to .90).

Characteristics Associated with Child Fatness

In the least squares models (Table 3), arm fat area and body mass index were examined as unadjusted raw variables; therefore, most of the explained variation was attributable to biological variables (R^2s for the biological models alone were 25.4% and 11.4% for arm fat area and body mass index, respectively, compared with 28.5% and 12.2% for the full models). As expected, fatness increased significantly with grade (age) and height, girls had higher arm fat area indices than boys,[22] and African Americans had lower arm fat area indices than Whites.[23]

In the arm fat area model, all sociodemographic variables entered except maternal employment were related to child fatness after adjustment for the others. In the body mass index model, similar results were found, although some relationships were slightly less significant.

TABLE 3—Multiple Least Squares Regression Models of Child Fatness as a Continuous Variable

Explanatory Variable	Dependent Variable: Arm Fat Area (n = 1135) Regression Coefficient	SE	Dependent Variable: Body Mass Index (n = 1138) Regression Coefficient	SE
Biological variables				
Grade (5th vs 2nd)	.36*	.030	.10*	.010
Age within grade, m[a]	−.007*	.003	.003*	.001
Height within grade, cm[a]	.046*	.005	. . .[b]	. . .
Sex (female vs male)	.28*	.058	.034	.019
Race (African American vs White)	−.32*	.062	−.040	.021
Sociodemographic variables				
SES score (medium/high vs low)	−.094	.052	−.031	.018
Family structure (single parent vs dual parent)	−.15*	.062	−.046*	.021
School lunch (participants vs nonparticipants)	.12*	.037	.039*	.012
Siblings 1 (single child vs 1 or more siblings)	.10*	.046	.020	.015
Siblings 2 (3 or more siblings vs 2 or less)	−.084*	.042	−.027	.014
Breakfast skipping (no breakfast vs some breakfast)	.11*	.056	.063*	.019
Interactions				
SES × Family Structure	.22*	.084	.072*	.028
SES × Height	−.014*	.005	. . .	. . .
SES × Sex	−.10	.067	−.032	.022
Breakfast × Family Structure	−.26*	.12	−.14*	.038
Intercept	6.76*	.060	2.81*	.020

Note. SES = socioeconomic status.
[a]Centered around the mean for each grade.
[b]Not entered (already adjusted for in dependent variable).
*$P < .05$.

Low-SES children tended to be fatter (by about 5%) than medium/high-SES children, except in single-parent households (an interaction). The inverse relationship of fatness with SES was stronger in girls than in boys; although this interaction was slightly less significant than the $P < .10$ cutoff, it was kept in the model because of the similar findings in adults.

Children in single-parent families tended to be thinner than those in two-parent families. Children who participated in school lunch tended to be slightly but significantly fatter than those who did not; this relationship held even when children eligible for free or reduced-price lunches were omitted from the comparison. Children with no or few siblings tended to be fatter than those with many siblings. Finally, children who skipped breakfast tended to be fatter than those who ate breakfast, except in single-parent households. The other dietary indices examined did not relate to child fatness.

On the basis of the body mass index model, these relationships translate into low-SES children being from 0.4 to 1.8 kg (1 to 4 lb) heavier (depending on their other characteristics), single-parent children being up to 0.4 kg (1 lb) lighter, school lunch participants being 0.7 to 0.9 kg (1.5 to 2 lb) heavier, children with no siblings being 0.4 to 0.7 kg (1 to 1.5 lb) heavier, and children with three or more siblings being 0.4 to 0.7 kg (1 to 1.5 lb) lighter than their respective counterparts.

Characteristics Associated with Overweight and Underweight

On the basis of chi-square tests with no adjustment for other variables, there was significantly more overweight (> 90th percentile for body mass index or arm fat area) in children with few or no siblings, those with employed mothers, and those who ate school lunch. There was more underweight (≤ 10th percentile) in children with single parents.

The logistic regression models (Table 4) mostly supported the findings of the least squares models. African-American children were slightly less likely to be overweight, based on body mass index, and much more likely to be "underfat," based on arm fat area, than White children. In two-parent households, medium/high-SES children were about half as likely as low-SES children to be overweight and slightly less likely to be underweight. In single-parent households, on the other hand, SES was not significantly related to overweight or underweight, except for a trend toward medium/high-SES children being heavier than low-SES children.

Compared with children who did not eat school lunch, children who ate school lunch were about half as likely to be underweight and somewhat, but not significantly, more likely to be overweight. Children with no siblings were somewhat more likely, and those with three or more siblings significantly less likely, to be overweight than those with one or two siblings. In two-parent families, children who skipped breakfast were almost twice as likely as breakfast eaters to be overweight; they were less likely to be underweight, as measured by arm fat area, than those who ate breakfast.

Discussion

Because the exact degree of child fatness related to later health problems is unknown, there is no health-based definition of child overweight. Its diagnosis thus remains statistical rather than pathological, and the prevalence in a given population depends on the reference data, the cutoff level, and the indicator used.

The most useful reference data are those representative of the US population. Therefore, the sample measurements were compared with the national NHANES I and NHANES II survey data. Because children in the nation as a whole became fatter during the 1970s,[2] the prevalence of overweight in the sample is higher when compared with NHANES I national percentiles than with NHANES II percentiles, even though both data sets are representative samples of healthy individuals in the United States.

As far as cutoff level, values over the 95th percentile clearly represent a nutritional concern. Many also recommend use of the 85th percentile to define overweight in children, as is used increasingly in adults.[24–26] All three of the fatness indicators resulted in similar prevalences

TABLE 4—Odds Ratios (95% Confidence Intervals) of Being Overweight or Underweight Compared with Normal Weight, Based on Logistic Regression, for Each Explanatory Characteristic

Explanatory Characteristic	Dependent Variables: Overweight (>90th Percentile) Body Mass Index	Overweight (>90th Percentile) Arm Fat Area	Underweight (≤10th Percentile) Body Mass Index	Underweight (≤10th Percentile) Arm Fat Area
African American vs White	0.43[a] (0.19, 0.98)	0.87 (0.44, 1.73)	1.57 (0.72, 3.43)	3.63[a] (1.94, 6.80)
Medium/high vs low SES[b]				
In 2-parent families	0.53[a] (0.35, 0.80)	0.59[a] (0.40, 0.87)	0.57 (0.30, 1.08)	0.79 (0.42, 1.47)
In 1-parent families	2.01 (0.93, 4.36)	1.14 (0.53, 2.43)	0.76 (0.26, 2.25)	0.72 (0.28, 1.83)
Single-parent vs dual-parent family[b]				
In low-SES families	0.54 (0.27, 1.08)	0.62 (0.33, 1.16)	0.90 (0.38, 2.12)	1.28 (0.59, 2.81)
In medium/high-SES families	2.05[a] (1.12, 3.77)	1.20 (0.66, 2.18)	1.21 (0.48, 3.07)	1.17 (0.52, 2.63)
School lunch participants vs nonparticipants	1.48 (0.96, 2.27)	1.40 (0.95, 2.08)	0.66 (0.38, 1.14)	0.51[a] (0.32, 0.82)
Single child vs 1–2 siblings	1.46 (0.91, 2.34)	1.35 (0.88, 2.08)	1.51 (0.79, 2.88)	0.96 (0.50, 1.83)
3 or more siblings vs 1–2	0.60[a] (0.37, 0.98)	0.50[a] (0.31, 0.82)	0.66 (0.32, 1.37)	0.90 (0.50, 1.61)
No breakfast vs some[b]				
In 2-parent families	1.95[a] (1.17, 3.25)	1.51 (0.92, 2.46)	1.36 (0.59, 3.17)	0.23[a] (0.06, 0.97)
In 1-parent families	0.64 (0.20, 2.03)	0.43 (0.12, 1.54)	0.69 (0.15, 3.28)	0.61 (0.17, 2.22)

Note. The ratios for each explanatory characteristic have been adjusted for the other characteristics. SES = socioeconomic status.
[a]Significant based on 95% confidence interval not including odds ratio of 1.0.
[b]Significant interaction with following variable.

of overweight, although body mass index tended to give higher prevalences than triceps skinfold or arm fat area.

Regardless of the definition, the prevalence of overweight in the sample is substantially higher than that in the nation in 1974 or 1980. This is a concern because it may result in increased rates of adult obesity and rising health care costs, supporting the need for obesity prevention efforts starting in childhood. While the results probably reflect a greater current prevalence of child overweight in New York State than in the nation overall, similar to that found in the Northeast subsample of NHANES II, children nationally may well have increased in fatness between the most recent national survey and the present New York survey.

The prevalence of underweight in the study sample was similar to that in the nation in 1980, suggesting less of a problem than for overweight. However, this finding does not imply adequate dietary quality or the absence of at least episodic hunger (food insecurity) or that some children are not experiencing health problems due to being underweight.

The multivariate analyses identified several factors related to the degree of fatness in children. The lower arm fat area but not body mass index values of African Americans relative to Whites, similar to skinfold vs body weight findings from other studies,[22,26] may relate to racial differences in fat patterning rather than to differences in overall fatness.

The relationship between SES and child overweight is unclear in the literature. An inverse relationship like that revealed in this study was found in three other large US studies, including the National Health Examination Survey.[27–29] However, a positive relationship was found in the Ten State Nutrition Survey (except for an inverse relationship in late teen girls)[30] and the Tecumseh community study.[31] No relationship was found in three other large US studies.[32–34] In an extensive review of studies of all sizes in the United States and other developed countries, no clear trend emerged.[35] However, most of the relationships in girls were in the inverse direction, similar to the trend found in the present study and corresponding to differences found in adults.[35] This suggests that the trend toward females of lower SES being fatter may begin in childhood.

The interaction between SES and family structure, such that low-SES children tended to be fatter than higher SES children except in single-parent households, may be due to single-parent low-SES families being even poorer (in all necessary resources) than two-parent low-SES families. Although further study and better measures of SES are needed, this suggests a nonlinear relationship between SES and overweight, such that overweight is lowest in the very poor and in the wealthy and highest in the somewhat poor. Of four other studies that examined the relationship between child overweight and single parenthood, two showed a positive association[36,37] and two showed no association[38,39]; however, none controlled for SES.

A positive relationship between child fatness and participation in school lunch also was found in the national evaluation of the School Lunch Program.[40] The causal nature of this relationship is unclear, however. For example, children who choose not to eat school lunch could be leaner because they are pickier eaters or have more health-conscious parents. On the other hand, school lunches have been suggested to be high in fat,[41–43] which could lead to greater fatness in children.

The inverse relationship between the number of children in the family and child fatness has been found in a number of other studies,[36,44,45] but the mechanism remains unclear. Children with more siblings may be more physically active.

The positive association between breakfast skipping and overweight also was found in a study of adolescents,[46] but the pathway is unclear. Those who skip breakfast may eat more later in the day or be less active, or already-overweight children may tend to skip breakfast in an attempt to lose weight.

Maternal employment was not related to child overweight, as found elsewhere.[38,44] Other factors reported to relate to child overweight but not examined here include activity level and television viewing.[47]

In conclusion, the prevalence of overweight among school-aged children in New York State outside of New York City (14% to 28%, depending on the definition) is high in comparison with national reference data. Children who tended to be fatter were members of low-SES, two-parent households, those with few or no siblings, those who ate school lunch, and those who skipped breakfast. These variables may be useful for targeting interventions to prevent child overweight, although further research is needed to better understand the relationships and their mechanisms and to identify other related factors. □

Acknowledgments

Support for this work was provided by the Nutrition Surveillance Program of the New York State Department of Health and a National Institutes of Health nutrition training grant (DK07158).

We would like to thank Dr Jay Kumar and Annette Palk of the New York State Dental Bureau for sample selection and recruitment, Dr Tom Santner and Gail Rothe for help with data analysis, Janet Weber for the initial data collection, Dr Mary Fran Sowers and Deb Spicer for help in the initial design of the project, and Gayle Leitch Kelly for her administrative support.

References

1. Dietz WH. Childhood obesity: susceptibility, cause, and management. *J Pediatr.* 1983;103:676–686.
2. Gortmaker SL, Dietz WH, Sobol AM, Wehler CA. Increasing pediatric obesity in the United States. *Am J Dis Child.* 1987;141: 535–540.
3. Ross JA, Gilbert G. The National Children and Youth Fitness Study: a summary of findings. *J Phys Educ Recreation Dance.* 1985;56(suppl 1):45–50.
4. Rimm IJ, Rimm A. Association between juvenile onset of obesity and severe adult obesity in 73,532 women. *Am J Public Health.* 1976;66:479–489.
5. Garn SM, LaVelle M. Two-decade follow-up of fatness in early childhood. *Am J Dis Child.* 1985;139:181–185.
6. Cronk CE, Roche AF, Chumlea WC, Kent R. Longitudinal trends of weight/stature^2 in childhood in relationship to adulthood body fat measures. *Hum Biol.* 1982;54:751–764.
7. Abraham S, Nordsieck M. Relationship of excess weight in children and adults. *Public Health Rep.* 1960;75:263–273.
8. Dietz WH. Obesity in infants, children, and adolescents in the United States, I. Identification, natural history, and aftereffects. *Nutr Res.* 1981;1:117–137.
9. Wolfe WS, Campbell CC. Food pattern, diet quality and related characteristics of school children in New York State. *J Am Diet Assoc.* 1993;93:1280–1284.
10. Kumar J, Green E, Wallace W, Bustard R. Changes in dental caries prevalence in upstate New York school children. *J Public Health Dent.* 1991;51:158–163.
11. Snedecor GW, Cochran WG. *Statistical Methods.* 7th ed. Ames, Iowa: Iowa State University Press; 1980.
12. Cole TJ. A critique of the NCHS weight for height standard. *Hum Biol.* 1985;57:183–196.
13. Roche AF, Siervogel RM, Chumlea WC, Webb P. Grading body fatness from limited anthropometric data. *Am J Clin Nutr.* 1981;34:2831–2838.
14. Frisancho AR. New norms of upper limb fat and muscle areas for assessment of nutritional status. *Am J Clin Nutr.* 1981;34: 2540–2545.
15. Anthropometric reference data and prevalence of overweight, United States, 1976–80. *Vital Health Stat [11].* 1987;238. DHHS publication PHS 87-1688.
16. Marks GC, Habicht J-P, Mueller WH. Reliability, dependability, and precision of anthropometric measurements: the Second National Health and Nutrition Examination Survey 1976–1980. *Am J Epidemiol.* 1989;130:578–587.
17. Dietz WH, Gortmaker SL. Factors within the physical environment associated with childhood obesity. *Am J Clin Nutr.* 1984;39: 619–624.
18. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research.* Belmont, Calif: Lifetime Learning Publications; 1982.
19. *SAS User's Guide: Statistics, Version 5 Edition.* Cary, NC: SAS Institute Inc; 1985.
20. Must A, Dallal GE, Dietz WH. Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht^2) and triceps skinfold thickness. *Am J Clin Nutr.* 1991;53: 839–844.
21. Hamill PV, Drizd TA, Johnson CL, et al. Physical growth: National Center for Health Statistics percentiles. *Am J Clin Nutr.* 1979;32:607–629.
22. Tanner JM. *Foetus into Man: Physical Growth from Conception to Maturity.* Cambridge, Mass: Harvard University Press; 1978.
23. Harsha DW, Voors AW, Berenson GS. Racial differences in subcutaneous fat pattern in children aged 7–15 years. *Am J Phys Anthropol.* 1980;53:333–337.
24. *The Surgeon General's Report on Nutrition and Health.* Washington, DC: US Dept of Health and Human Services; 1988. DHHS publication PHS 88-50210.
25. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: Govt Printing Office, US Dept of Health and Human Services; 1990. DHHS publication PHS 91-50212.
26. National Research Council, Committee on Diet and Health. *Diet and Health: Implications for Reducing Chronic Disease Risk.* Washington, DC: National Academy Press; 1989.
27. *Determinants of Children's Health.* Washington, DC: National Center for Health Services Research; 1981. DHHS publication PHS 81-3309.
28. Stunkard AJ, D'Aquili E, Fox S, Filion RDL. The influence of social class on obesity and thinness in children. *JAMA.* 1972;221:579–584.
29. Huenemann RL, Shapiro LR, Hampton MC, Mitchell MW. A longitudinal study of gross body composition and body conformation and their association with food and activity in a teenage population. *Am J Clin Nutr.* 1966;18:325–338.
30. Garn SM, Clark DC. Nutrition, growth, development and maturation: findings of the Ten State Nutrition Survey of 1968–70. *Pediatrics.* 1975;56:306–319.
31. Garn SM, Hopkins PJ, Ryan AS. Differential fatness gain of low income boys and girls. *Am J Clin Nutr.* 1981;34:1465–1468.
32. Khoury PR, Morrison JA, Laskarzenski P, et al. Relationships of education and occupation to coronary heart risk factors in schoolchildren and adults: the Princeton School District study. *Am J Epidemiol.* 1981;113:378–395.
33. Hunter S, Frerichs PR, Webber LS, Berenson GS. Social status and cardiovascular disease risk factor variables in children: the Bogalusa Heart Study. *J Chronic Dis.* 1979;32:441–449.
34. Malina RM. Patterns of development in skinfolds of Negro and White Philadelphia children. *Hum Biol.* 1966;38:89–103.
35. Sobal J, Stunkard AJ. Socioeconomic status and obesity: a review of the literature. *Psychol Bull.* 1989;105:260–275.
36. Jacoby A, Altman DG, Cook J, Holland WW. Influence of some social and environmental factors on the nutrient intake and nutritional status of schoolchildren. *Br J Prev Soc Med.* 1975;29:116–120.
37. Garman AR, Chinn S, Rona RJ. Comparative growth of primary schoolchildren from one and two parent families. *Arch Dis Child.* 1982;57:453–458.
38. Bowering J, Wynn RL. Nutritional status of preschool children from intact and divorced families. *Home Econ Res J.* 1986; 15:132–140.
39. Darke SJ, Disselduff MM, Try GP. A nutrition survey of children from one-parent families in Newcastle upon Tyne in 1970. *Br J Nutr.* 1980;44:237–241.
40. Vermeersch J, Hanes S, Gale S. The national evaluation of school nutrition programs: program impact on anthropometric measures. *Am J Clin Nutr.* 1984;40: 414–424.
41. Dillon MS, Lane HW. Evaluation of the offer vs. serve option within self-serve, choice menu lunch program at the elementary school level. *J Am Diet Assoc.* 1989;89: 1780–1785.
42. Coale EK, Bedford MR. Fat-controlled menus in a school lunch program. *Sch Food Serv Res Rev.* 1984;8:37–41.
43. Morris PM. *Heading for a Health Crisis: Eating Patterns of America's School Children.* Washington, DC: Public Voice for Food and Health Policy; 1991.

44. Rona RJ, Chinn S. National Study of Health and Growth: social and biological factors associated with weight-for-height and triceps skinfold of children from ethnic groups in England. *Ann Hum Biol.* 1987;14: 231–248.
45. Samuelson G, Grahnen H, Arvidsson E. An epidemiological study of child health and nutrition in a northern Swedish county, VI: relationship between general and oral health, food habits, and socioeconomic conditions. *Am J Clin-Nutr.* 1971;24:1361–1373.
46. Resnicow K. The relationship between breakfast habits and plasma cholesterol levels in schoolchildren. *J Sch Health.* 1991;61:81–85.
47. Gortmaker SL, Dietz WH, Cheung LWY. Inactivity, diet, and the fattening of America. *J Am Diet Assoc.* 1990;90:1247–1255.

5-A-DAY: Dietary Behavior and the Fruit and Vegetable Intake of Latino Children

Charles E. Basch, PhD, Patricia Zybert, PhD, and Steven Shea, MD

ABSTRACT

Objectives. The purpose of the study was to examine children's intake of fruits and vegetables in relation to the recent national "5-A-DAY" campaign.

Methods. Four 24-hour dietary recalls per child collected from 205 mothers of 4- to 5-year-old urban Latino children were used to analyze average 5-A-DAY fruit and vegetable consumption and examine associations between 5-A-DAY consumption, nutrient intakes, and eating patterns.

Results. The reported mean servings per day of fruits and vegetables, as defined by 5-A-DAY criteria, were 1.8 and 1.0, respectively, with only 6.8% (n = 14) of the children averaging five or more servings per day. Fruit juice accounted for 36% of 5-A-DAY servings. There were significant linear trends in intake of vitamins A and C, potassium, iron, cholesterol, protein, and fiber across quintiles of 5-A-DAY intake. There were no differences among quintiles in intake of saturated or total fat or in servings from most non–5-A-DAY food groups.

Conclusions. Latino children's intake of fruits and vegetables falls far short of current recommendations. Fruit juice accounted for a disproportionate amount of 5-A-DAY intake in this population. Sensible 5-A-DAY interventions should take into consideration the existing eating patterns of the target population. (*Am J Public Health.* 1994;84: 814–818)

Introduction

Systematic reviews of epidemiologic studies in adults indicate that higher levels of fruit and vegetable intake are associated with reduced risk of cancer at several sites.[1–8] As a result, as one of its cancer and nutrition risk reduction objectives for the year 2000, the US Public Health Service,[1] with wide support from other agencies,[9–13] has called for adults to increase servings of fruits and vegetables from an estimated current average of 2.8 servings per day to five or more servings per day. In accordance with this goal, the National Cancer Institute (NCI), with the cooperation of selected food industry representatives, has sponsored a multifaceted national campaign known as "5-A-DAY."[14] Although the 5-A-DAY recommendation was formulated with adults in mind, the Public Health Service has also recommended that children 2 years of age and older progress toward this dietary pattern.[1]

The weight of existing evidence supports protective health effects for a dietary pattern high in fruit and vegetable intake without identifying the specific vitamins, minerals, or nutrients responsible for these effects.[4,5,15] Causal mechanisms involved in the relationship between fruit and vegetable consumption and reduced cancer risk[4] may involve individual responses to dietary factors and nutrient interactions, which are themselves determined by genetic, physiologic, and life-style factors.[15] Thus, even though basic and epidemiologic research have identified specific vitamins, minerals, micronutrients, and macronutrients as the most likely agents responsible for the association between fruit and vegetable intake and cancer, the 5-A-DAY message is not to increase intake of these nutrients per se but to increase intake of fruits and vegetables. The emphasis on foods as opposed to nutrients[16] exemplified by the 5-A-DAY program reflects a relatively recent focus in public health nutrition.

While there has been some research describing fruit and vegetable consumption patterns and associated nutrient intakes in adults,[17,18] we did not identify any such studies in young children. Current eating behaviors of a target population should guide the formulation of effective, acceptable, and sustainable 5-A-DAY intervention strategies. We therefore analyzed repeated 24-hour dietary recalls in a sample of 205 urban Latino children to describe intake of 5-A-DAY foods, to determine whether high 5-A-DAY food consumption was associated with differential intake of specific nutrients and calories, and to examine whether high 5-A-DAY consumers substituted fruits and vegetables for other foods not in the 5-A-DAY group or added fruits and vegetables to a similar underlying diet.

Methods

Study Sample

Subjects were young healthy Latino children and their mothers (n = 214 mother/child pairs) who participated in a 3-year observational study of children's

Charles E. Basch and Patricia Zybert are with the Center for Health Promotion, Teachers College, Columbia University, New York City. Steven Shea is with the College of Physicians and Surgeons, Columbia University, and the Division of Epidemiology, Columbia University School of Public Health.

Requests for reprints should be sent to Charles E. Basch, PhD, Center for Health Promotion, Box 114, Teachers College, Columbia University, New York, NY 10027.

This paper was accepted January 31, 1994.

Editor's Note. See related editorial by Nestle (p 713) in this issue.

TABLE 1—Mean Number of 5-A-DAY Servings per Day Derived from Four 24-Hour Dietary Recalls, by Quintiles of 5-A-DAY Servings

	Lowest Quintile (n = 41)		Second Lowest Quintile (n = 41)		Middle Quintile (n = 41)		Second Highest Quintile (n = 41)		Highest Quintile (n = 41)		Total Sample (n = 205)		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Median
All 5-A-DAY foods	1.3	0.4	2.1	0.2	2.7	0.2	3.3	0.2	4.8	0.9*	2.8	1.3	2.7
All 5-A-DAY fruits	0.9	0.4	1.3	0.5	1.7	0.5	2.1	0.5	3.0	1.0*	1.8	0.9	1.8
Fresh fruits	0.2	0.3	0.4	0.3	0.5	0.4	0.6	0.4	0.9	0.6*	0.5	0.5	0.4
Other fruits	0.1	0.2	0.3	0.3	0.3	0.4	0.3	0.3	0.3	0.4	0.3	0.3	0.3
Juices	0.5	0.4	0.7	0.4	0.9	0.4	1.2	0.5	1.7	0.9*	1.0	0.7	0.9
Orange juice	0.4	0.4	0.5	0.4	0.7	0.5	0.9	0.5	1.3	0.8	0.8	0.6	0.7
Apple juice	0.1	0.2	0.1	0.2	0.2	0.3	0.2	0.2	0.4	0.6	0.2	0.4	0.0
All 5-A-DAY vegetables	0.5	0.3	0.8	0.5	0.9	0.5	1.2	0.5	1.8	0.8*	1.0	0.7	1.0
Dark green/yellow vegetables	0.0	0.0	0.1	0.2	0.1	0.2	0.2	0.3	0.2	0.2*	0.1	0.2	0.0
Beans	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.3	0.6	0.4	0.4	0.3	0.3
Other vegetables	0.2	0.3	0.4	0.5	0.5	0.4	0.6	0.4	1.1	0.6*	0.6	0.5	0.4

*$P < .001$ (linear trend).

diet and physical activity. All subjects were living in New York City. None of the children were on special diets. Recruitment was mainly from a large urban pediatric practice and has been described in greater detail elsewhere.[19] Analyses for the present study included all Latino mother/child pairs with four completed 24-hour dietary recalls for the child taken during the first year of the study (n = 205; 103 male children). At the time of the first recall, children's mean age was 4.4 years (SD = 0.34) and mothers' mean age was 31.8 years (SD = 5.2). Spanish was the language spoken at home by 95% (n = 195) of these families. More than 90% of the children for whom we had birthplace data (n = 184) were born in the United States. Seventy-four percent of the mothers were born in the Dominican Republic. Ninety percent of the mothers received Medicaid.

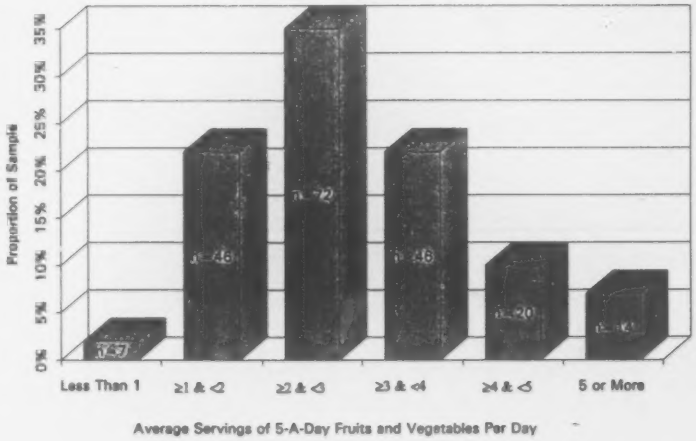


FIGURE 1—Frequency distribution for proportion of children (n = 205) who consumed varying serving levels of 5-A-DAY fruits and vegetables derived from an average of four dietary recalls over 1 year.

Measurements

From 1986 to 1989, trained research staff interviewed mothers (usually in Spanish) at our community-based field site. An existing reliable and validated protocol[20,21] was used to collect up to seven 24-hour dietary recalls over the course of 3 years. Additional descriptions of the data collection procedures and the nutrient analysis program and database have appeared in other published reports, and support for the reliability and validity of mothers' reports of children's diet has been shown.[20,21]

Typical nutrient intakes for each child were derived by averaging results from the four 24-hour recalls collected during the first year of the study. Inclusion/exclusion criteria and serving size guidelines specified by NCI[14] were used to calculate the number of servings of 5-A-DAY fruits and vegetables for each recall. Avocados, coconut, olives, nuts, and seeds were excluded, as were foods that were sweetened or fried. Only 100% nonsweetened fruit juices were included. Lacking complete information in our dietary recalls and nutrient database, we assumed that all apple juice consumed was unsweetened. Mixed vegetable dishes were included if the percentage of calories from fat was less than 30, the percentage of calories from saturated fat was less than 10, and the number of milligrams of sodium per serving was less than 360.[14] 5-A-DAY standard serving sizes are one piece of fresh fruit, one-half cup of cooked fruit and vegetables, six fluid ounces of juice, one cup of leafy vegetables, or one-fourth cup of dried fruit. In this paper, 5-A-DAY fruit and vegetable consumption represents the sum

TABLE 2—Daily Intake of Selected Nutrients and Total Calories Derived from Four 24-Hour Dietary Recalls, by Quintiles of 5-A-DAY Servings

	Lowest Quintile (n = 41)		Second Lowest Quintile (n = 41)		Middle Quintile (n = 41)		Second Highest Quintile (n = 41)		Highest Quintile (n = 41)		Total Sample (n = 205)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total energy intake, calories	1484	348	1482	297	1536	356	1545	376	1690	403*	1547	362
Protein, g	55.0	18.5	54.7	12.2	56.7	16.9	60.7	13.6	64.9	14.5**	58.4	15.7
Carbohydrates, g	190.3	41.3	197.4	42.1	205.4	44.0	202.6	49.3	224.4	51.8*	204.0	46.9
Dietary fiber, g	8.0	3.6	9.1	3.0	9.8	2.2	12.0	3.6	14.8	4.4**	10.7	4.2
Saturated fat, g	22.5	7.1	22.4	6.0	22.2	6.8	23.6	7.3	24.1	8.2	23.0	7.1
Saturated fat, % of calories	13.7	2.7	13.6	2.2	12.9	2.1	13.6	2.0	12.7	2.0	13.3	2.2
Total fat, g	56.4	19.0	53.6	14.5	55.4	16.6	55.9	17.4	60.6	19.1	56.4	17.4
Total fat, % of calories	33.8	5.0	32.4	4.3	32.1	3.8	31.2	3.3	32.0	3.8	32.5	4.1
Cholesterol, mg	248.1	105.7	295.7	145.6	290.7	121.3	313.2	120.0	362.3	125.8**	302.0	128.6
Cholesterol/1000 kcal	168.6	60.0	200.1	83.1	189.4	69.2	203.4	67.8	216.2	62.4*	195.5	70.1
Vitamin A, IU	3227	3115	3494	2074	3920	1906	5503	3339	5594	2981**	4348	2899
Vitamin C, mg	90.9	36.0	100.8	39.1	115.5	33.4	123.6	29.1	155.7	58.5**	117.3	45.9
Calcium, mg	818.3	290.8	820.0	200.8	815.0	294.7	904.6	215.1	931.8	295.9	857.9	265.1
Potassium, mg	1928	469	2034	294	2201	466	2444	437	2859	563**	2293	560
Iron, mg	8.5	3.0	8.6	2.6	9.6	2.8	10.0	2.7	11.2	4.3**	9.6	3.3

*$P < .01$ (linear trend); **$P < .001$ (linear trend).

across all reported eating occasions of 5-A-DAY eligible fruits and vegetables presented in terms of 5-A-DAY standard serving sizes.

5-A-DAY fruits were categorized as (1) fresh fruits, (2) fruit juices, or (3) others. 5-A-DAY vegetables were categorized as (1) beans and legumes; (2) dark green leafy, green nonleafy, or deep yellow vegetables; or (3) others.[14] This last vegetable group included homemade stews and soups meeting the NCI sodium and fat restrictions if one cup of the recipe provided at least one-half cup of vegetables.*

Statistical Analysis

Means and standard deviations for typical daily 5-A-DAY fruit and vegetable consumption were calculated. Children were then classified into quintiles of 5-A-DAY servings per day. Linear trends[22] across quintiles were tested for calories and selected nutrients and for numbers of servings from each food group. The regression technique recommended by Willett[23–25] was used to adjust nutrients to control for caloric intake. The natural logarithm transformation [ln $(x + 0.5)$] was used to normalize skewed distributions of both nutrient intakes and numbers of servings. Untransformed values are presented in the tables, but all statistical analyses involving significance testing and calorie adjustment of nutrients were performed on transformed data.

Results

The mean number of servings of 5-A-DAY fruits and vegetables consumed per day was 2.8 (SD = 1.3), and the median was 2.7 (Table 1). There was no significant difference in the proportion of boys and girls across the quintiles of 5-A-DAY intake ($\chi^2 = 2.7$, 4 *df*). There was substantial variability in consumption of fruits and vegetables across quintiles. In comparison with children in the lowest quintile, children in the highest quintile consumed more than three and a half times as many 5-A-DAY servings (4.8 vs 1.3) (Table 1). Only 14 of the 205 children in the study (6.8%) averaged the five or more servings a day recommended by the Public Health Service[1] (Figure 1).

About two thirds of the children (n = 128) consumed less than two servings of fruit per day. Only one child consumed three or more servings of vegetables per day. Children in all quintiles consumed more servings of fruit than of vegetables (mean = 1.8 vs 1.0 for the total sample) (paired $t = 10.7$, $df = 204$, $P < .001$). Fruit juice was the largest source of 5-A-DAY servings, accounting for 36% of total 5-A-DAY intake and 56% of 5-A-DAY fruit intake for the sample as a whole. Orange juice was by far the most popular fruit consumed in our sample. Dark green leafy, green nonleafy, and deep yellow vegetables were consumed least frequently (mean = 0.1, SD = 0.2, for the total sample).

The children who consumed more 5-A-DAY servings per day had slightly higher daily caloric intakes ($P < .01$), with a difference of approximately 14% (206 calories) between the lowest and highest quintiles of 5-A-DAY servings (Table 2). There were significant linear trends ($P < .001$) in intake of vitamins A and C, potassium, iron, cholesterol, protein and fiber, with intakes increasing from lowest to highest quintiles of 5-A-DAY intake. After calorie adjustment, the significance of the linear trends for cholesterol and protein dropped to $P < .01$. The others remained unchanged. There were no differences across quintiles in percentage of calories from saturated or total fat.

With regard to intake of non–5-A-DAY foods, children in the highest quintile of 5-A-DAY food intake ate more than twice as many eggs as those in the lowest quintile (0.7 vs 0.3 eggs per day), explaining their higher dietary cholesterol intake. Children in the highest 5-A-DAY quintile also consumed fewer servings of fruit punch and soda (0.3 vs 0.8 servings per day) and slightly more un-

*Appendixes ("Appendix A—Mean No. of 5-A-DAY Servings per Eating Occasion" and "Appendix B—Thirty-three Food Groups [US Dept of Agriculture Typology] Excluding 5-A-DAY Fruits and Vegetables") are available from the author.

fried poultry and fried fish. There were no other differences across the quintiles in numbers of servings from any of the non–5-A-DAY food groups analyzed.

Discussion

In our sample of 4- to 5-year-old urban Latino children, even those in the highest quintile of 5-A-DAY fruit and vegetable consumption averaged fewer than the five servings a day currently recommended by the Public Health Service.[1] Only about one third averaged two servings of fruit per day, and just one child averaged three servings of vegetables per day. Fruit juice, usually orange juice, was the most prevalent source of 5-A-DAY servings. Children in our sample averaged one serving of fruit juice per day but only one half of one serving of fresh fruit per day. They consumed almost no dark green leafy, green nonleafy, or deep yellow vegetables.

It is perhaps unreasonable to expect young children to eat the same number of adult-sized portions of fruits and vegetables as adults. In our sample, the standardized 5-A-DAY fruit serving sizes were compatible with children's actual consumption patterns, but the standardized 5-A-DAY vegetable serving sizes were generally too large (data not shown). Nevertheless, even if we halve vegetable serving sizes, thereby doubling the number of servings of vegetables, only children in the top quintile of 5-A-DAY consumption average five or more servings per day.

The direct associations between servings of fruits and vegetables and intakes of vitamins A and C, potassium, and dietary fiber were expected. Fruits and vegetables are rich sources of these vitamins and nutrients. There was no evidence that implementation of the 5-A-DAY dietary guideline would reduce children's percentage of calories from saturated or total fat.

On the basis of analyses of variance of the average number of servings from each of the non–5-A-DAY food groups by 5-A-DAY classification, the children in our sample who ate more 5-A-DAY servings generally did not substitute fruits and vegetables for other foods (data not shown). They added fruits and vegetables to an underlying diet that did not differ greatly across quintiles of 5-A-DAY intake. The one substitution we found was fruit juice for fruit punch and soda. Children in the highest quintile consumed 3.4 times as much fruit juice a day as children in the lowest quintile but less than half the amount of punch and soda. Examining individual recalls rather than averages for each child, we found that punch and fruit juice were essentially interchangeable. Days with low juice consumption were characterized by high punch consumption, and vice versa (data not shown). The only other statistically significant differences found were higher egg, unfried poultry, and fried fish consumption for high 5-A-DAY consumers.

Patterson et al. analyzed dietary recall data from the second National Health and Nutrition Examination Survey (NHANES II), which represents the adult civilian noninstitutionalized population of the United States in 1976 through 1980.[18] Few adults (9%) consumed three or more servings of vegetables and two or more servings of fruit a day. Vegetables were more commonly eaten than fruits, with mean servings per day of 1.8 and 1.1, respectively. Among the children we studied, fruits were more commonly eaten than vegetables. This reversal suggests the need for different emphases in 5-A-DAY campaigns directed at children and at adults.

In the NHANES II data, total fat intake increased with servings of fruits and vegetables, and the percentage of calories from total and saturated fat varied directly with servings of vegetables and inversely with servings of fruits. In our data, total fat intake did not rise with increased 5-A-DAY consumption. When analyzing the data to determine whether intake of fruits alone or vegetables alone was associated with intake of percentage of calories from saturated and total fats, we found no association between percentage of calories from fats and fruit intake and an inverse association between percentage of calories from fats and vegetable intake (data not shown). These differences are not surprising when methods are compared. We followed the recently formulated NCI 5-A-DAY inclusion/exclusion criteria and serving size definitions,[14] which the Patterson paper predates.[18] Particularly relevant is the fact that we excluded fried vegetables (e.g., french fried potatoes), while Patterson et al. did not. French fries were mentioned 80 times in our 820 recalls, and the average amount consumed per eating occasion was 13 strips. Furthermore, Patterson's "servings" are what we would call "mentions." They are not presented in terms of a fixed, predefined serving size for each food. In neither study was there a relationship between servings of fruits and vegetables and body mass index (data not shown).

Several limitations of our study warrant mention. First, the nutrient database we used, which was based on US Department of Agriculture Handbook 7[26,27] and supplemented by other sources,[28–31] did not always permit us to distinguish between allowed and disallowed fruits and vegetables, the most relevant instance being sweetened vs unsweetened apple juice. Nevertheless, we believe that our recall data adequately represent children's dietary intakes from both behavioral and nutrient perspectives.[19–21] Second, while our data collection plan was not designed to assess directly the effects of seasonal variation, we did check for nonrandomness in the distribution of season of dietary recall by quintile of 5-A-DAY consumption. There was no relationship between season of recall and 5-A-DAY quintile. Third, we attempted to estimate typical consumption using only four 24-hour recalls over a 1-year period. In this sample, the intraclass correlation for the number of 5-A-DAY servings was .55 (.46 for fruits and .57 for vegetables). Finally, our assessment of children's diet was based on mothers' reports. Reanalyzing previously published data on the validity of such reports from a subsample of 55 children in this same study,[21] we found significant overreporting in only one of the six 5-A-DAY fruit and vegetable categories, namely dark green and yellow vegetables, and no significant underreporting. In the data analyzed here, children's average intake of dark green and yellow vegetables was only 0.1 servings per day. These findings support the reliability and validity of the data set as used in the context of this paper.

Generalizability of our results to groups of children of different ages or other ethnic backgrounds may be limited. The Latino population is important from a public health perspective because it is a substantial and rapidly growing segment of the population[32] that, relative to non-Hispanic Whites, has been shown to purchase[33] and consume[34] fewer low-fat foods. Our population was of low income (90% received Medicaid), and there is evidence in adults that those with lower incomes generally eat fewer fruits and vegetables.[17] Also, language and cultural differences may result in this population being harder to reach with public health nutrition campaigns aimed at the general public.

Sensible 5-A-DAY interventions must take into consideration the existing eating patterns of the target population. Fruit juices—particularly apple juice—

make relatively small contributions per calorie to intakes of vitamins A and C and dietary fiber, nutrients frequently implicated in the protective effect of a diet high in fruits and vegetables. For example, a 6-oz serving of frozen orange juice contains approximately 83 calories, 143 IU of vitamin A, 73 mg of vitamin C, and 0.8 g of fiber. The same amount of apple juice contains approximately 87 calories, no vitamin A, less than 2 mg of vitamin C, and 0.5 g of fiber. This issue of empty calories is especially relevant in our population since 44% of the children in our sample were above the 85th percentile of national norms for body mass index for their age and sex.[35] On the basis of our study, we would advocate increased consumption of fresh fruit and dark green leafy and nonleafy and deep yellow vegetables among children in our sample of low-income urban Latinos as a means of implementing the 5-A-DAY objectives. We would not advocate increased consumption of juice (especially fruit juice with a high calorie-to-nutrient ratio) in our sample. In addition, the Committee on Nutrition of the American Academy of Pediatrics[36] has expressed concern about gastrointestinal problems (chronic diarrhea, abdominal pain and bloating) related to excessive intakes of sorbitol from certain juices. The challenge for public health nutrition interventions is to find a balance between retaining the simplicity of the 5-A-DAY message and designing an intervention appropriate for the target population. □

Acknowledgments

This study was supported in part by grants from the National Heart, Lung, and Blood Institute (HL-35189 and HL-49508) and the Cancer Research Foundation of America.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1990. DHHS publication PHS-50212.
2. National Research Council. *Diet and Health: Implications for Reducing Chronic Disease Risk.* Washington, DC: National Academy Press; 1989.
3. *The Surgeon General's Report on Nutrition and Health 1988.* Washington, DC: US Dept of Health and Human Services; 1988.
4. Steinmetz KA, Potter JD. Vegetables, fruit, and cancer. I. Epidemiology. *Cancer Causes Control.* 1991;2:325–357.
5. Block G, Patterson B, Subar A. Fruit, vegetables, and cancer prevention: a review of the epidemiologic evidence. *Nutr Cancer.* 1992;18:1–29.
6. Byres T. Diet and cancer: any progress in the interim? *Cancer.* 1988;62:1713–1724.
7. Rogers AE, Longnecker MP. Biology of disease: dietary and nutritional influences on cancer: a review of epidemiological and experimental data. *Lab Invest.* 1988;59:729–759.
8. Henderson BE, Ross RU, Pike MC. Toward the primary prevention of cancer. *Science.* 1991;254:1131–1138.
9. *Cancer Control Objectives for the Nation: 1985–2000.* Washington, DC: National Cancer Institute; 1986. NIH publication 86-2880.
10. *Nutrition and Your Health: Dietary Guidelines for Americans.* Washington, DC: US Dept of Agriculture, US Dept of Health and Human Services; 1980. Home and Garden Bulletin 232.
11. National Research Council, Committee on Diet, Nutrition and Cancer. *Assembly of Life Sciences: Diet, Nutrition and Cancer.* Washington, DC: National Academy Press; 1982.
12. American Cancer Society. Nutrition and cancer: cause and prevention. Special report, American Cancer Society. *CA.* 1984;34:121–126.
13. Butrum RR, Clifford CK, Lanza E. NCI dietary guidelines: rationale. *Am J Clin Nutr.* 1988;48:888–895.
14. *5-A-DAY for Better Health.* Bethesda, Md: National Cancer Institute; 1992. RFA No. CA-92-17.
15. Council on Scientific Affairs, American Medical Association. Report of the Council on Scientific Affairs. Diet and cancer: where do matters stand? *Arch Intern Med.* 1993;153:50–56.
16. Willett WC. Challenges for public health nutrition in the 1990s. *Am J Public Health.* 1990;80:1295–1297.
17. Patterson BH, Block G. Food choices and the cancer guidelines. *Am J Public Health.* 1988;78:282–286.
18. Patterson BH, Block G, Rosenberger WF, Pee D, Kahle L. Fruit and vegetables in the American diet: data from NHANES II survey. *Am J Public Health.* 1990;80:1443–1449.
19. Shea S, Basch CE, Irigoyen M, et al. Relationships of dietary fat consumption to serum total and low density lipoprotein cholesterol in Hispanic preschool children. *Prev Med.* 1991;20:237–249.
20. Stein AD, Shea S, Basch CE, et al. Variability and tracking of nutrient intakes of preschool children based on multiple administrations of the 24-hour dietary recall. *Am J Epidemiol.* 1991;134:1427–1437.
21. Basch CE, Shea S, Arliss A, et al. Validation of mothers' reports of dietary intake by four to seven year-old children. *Am J Public Health.* 1990;80:1314–1317.
22. Snedecor GW, Cochran WG. *Statistical Methods.* 6th ed. Ames, Iowa: Iowa State University Press; 1967.
23. Willett WC. *Nutritional Epidemiology.* New York, NY: Oxford University Press; 1990.
24. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analysis. *Am J Epidemiol.* 1986;124:17–27.
25. Willett WC. Total energy intake and nutrient composition: dietary recommendations for epidemiologists. *Int J Cancer.* 1990;46:770–771.
26. Adams CF, Richardson M. *Nutritive Value of Foods.* Washington, DC: US Dept of Agriculture; 1981. Home and Garden Bulletin 72.
27. Gebhardt SE, Matthews RH. *Nutritive Value of Foods.* Washington, DC: US Dept of Agriculture; 1986. Home and Garden Bulletin 72.
28. de Reguero LC, de Santiago SMR. *Tabla de composicion de alimentos de uso corriente en Puerto Rico.* Río Piedras, Puerto Rico: Universidad de Puerto Rico; 1978.
29. Paul AA, Southgate DAT. *McCance and Widdowson's The Composition of Foods.* London, England: Her Majesty's Stationery Office; 1978.
30. Anderson JW. *Plant Fiber in Foods.* Lexington, Ky: HCF Diabetes Research Foundation; 1986.
31. Anderson JW, Bridges SR. Dietary fiber content of selected foods. *Am J Clin Nutr.* 1988;47:440–447.
32. US Department of Commerce, Bureau of Census; *The Hispanic Population in the United States: March 1988 Advance Report. Current Population Reports. Population Characteristics.* Series P-20, Washington, DC: US Dept of Commerce, Bureau of the Census; 1988. USGPO Publication 1988-201-478:80117.
33. Delapa RM, Mayer JA, Candelaria J, et al. Food purchase patterns in a Latino community: Project Salsa. *J Nutr Educ.* 1990;22:133–136.
34. Knapp JA, Hazuda HP, Haffner SM, et al. A saturated fat/cholesterol avoidance scale: sex and ethnic differences in a biethnic population. *J Am Diet Assoc.* 1988;88:172–177.
35. *Anthropometric Reference Data and Prevalence of Overweight: United States, 1976–80.* Hyattsville, Md: National Center for Health Statistics; 1987. DHHS publication PHS 87-1688.
36. Committee on Nutrition 1989–1990. *The Use of Fruit Juice in the Diets of Young Children.* Elk Grove Village, Ill: American Academy of Pediatrics; 1990.

Public Health Briefs

Family Networks: Predictors of Nursing Home Entry

Vicki A. Freedman, PhD, Lisa F. Berkman, PhD, Stephen R. Rapp, PhD, and Adrian M. Ostfeld, MD

A B S T R A C T

Despite the importance of kin in caring for older relatives, few studies have examined the relationship between the family network and the risk of nursing home placement. Data from a cohort of noninstitutionalized elderly persons living in New Haven in 1982 were used to predict nursing home admission through 1985. Logistic regression analyses show that older persons who have regular contact with kin have a lower risk of institutionalization. For men, the spouse is most important in reducing the risk of entry; for women, having regular contact with at least one family member—of any relation—reduces the risk of entry. (*Am J Public Health.* 1994;84:843–845)

Introduction

Each year, over 1 million elderly persons are admitted to nursing homes.[1] A number of prospective studies have examined measures related to nursing home admission.[2–15] The strongest and most consistent predictors of nursing home entry are functional status, mental status, and age. In addition, as first demonstrated by Townsend,[16] the absence of family ties appears to be associated with institutional residence. More recently, studies have documented that the living arrangements of an older person before entry,[4–6,11,12] the recent loss of a close tie,[4,6] and more general measures of social contact[14,15] predict nursing home admission.

Together these findings suggest that social networks—family members in particular—may be important in preventing the institutionalization of an older relative. Nevertheless, to date only limited measures of the social network have been considered in analyses of nursing home admission. In this study we explore the effects of both the composition and the size of the family network on the risk of institutionalization for elderly men and women.

Data and Methods

Data for this study are from the New Haven, Conn, site of the Established Populations for Epidemiologic Studies of the Elderly. The sampling design for the New Haven study has been described elsewhere.[17,18] Weighted estimates represent the population age 65 years and older in New Haven in 1982.

The New Haven cohort consisted of 2 812 noninstitutionalized men and women age 65 years and older in 1982. Respondents were interviewed annually from 1982 through 1991. Fewer than 1% of participants were lost to follow-up over the study period. Sample characteristics at baseline (1982) are presented in Table 1.

In this study, we used 3 years of follow-up data from 1982 to 1985, during which 354 participants experienced at least one nursing home episode. Respondents were asked annually about nursing home admissions. For each reported admission, the name of the nursing home and the approximate length of stay were recorded. Dates of entry and exit were then confirmed with named institutions.

Results

Table 2 presents the mean values for a series of family network variables, by gender. We included in our definition of family network children and other close relatives with whom the respondent had weekly or monthly face-to-face contact. Respondents identified close relatives in response to the question, "Apart from your children, how many other relatives do you have that you feel close to? (People you feel at ease with, can talk to about private matters, can call on for help?)" The spouse was also considered a

At the time of the study, Vicki A. Freedman, Lisa F. Berkman, and Adrian M. Ostfeld were with the Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Conn. Stephen R. Rapp is with the Department of Psychiatry and Behavioral Medicine, Bowman Gray School of Medicine, Wake Forest University, Winston-Salem, NC.

Requests for reprints should be sent to Vicki A. Freedman, PhD, Agency for Health Care Policy and Research, 2101 E Jefferson St, Suite 500, Rockville, MD 20852.

This paper was accepted August 3, 1993.

TABLE 1—Demographic and Health-Related Characteristics, New Haven EPESE, 1982

	Men, %	Women, %
Age group, y		
65–74	63.2	55.8
75–84	28.6	34.4
85+	8.3	9.7
Race		
White	82.6	81.6
Non-White	15.0	16.0
Missing	3.3	2.4
Education		
Less than high school	61.7	64.7
High school or more	37.8	24.8
Missing	2.2	2.6
Annual income, $		
< 5 000	17.8	35.5
5 000–10 000	37.1	32.1
> 10 000	34.6	17.9
Missing	10.5	14.5
Housing stratum		
Public low-income housing[a]	9.6	10.8
Private housing[b]	10.1	16.7
Community housing	80.3	72.5
Functional status[c]		
High (no limitations)	87.8	85.4
Low (1+ limitations)	11.3	13.9
Missing	0.9	0.6
Cognitive status[d]		
High (0–4 incorrect)	89.4	86.1
Low (4+ incorrect)	7.5	11.7
Missing	3.2	2.2
One or more nursing home admissions, 1982–1985	6.4	11.5
Cohort size (unweighted no.)	1 169	1 643

Note. Data are weighted. EPESE = Established Populations for Epidemiologic Studies of the Elderly.
[a]Housing is age and income restricted.
[b]Housing is age restricted.
[c]One point given for each activity for which respondent reported needed help: walking, bathing, grooming, dressing, eating, getting from bed to a chair, and using the toilet.
[d]One point given for each Pfeiffer mental status question respondent answered incorrectly.

TABLE 2—Family Network Composition and Size, New Haven EPESE, 1982

	% of Elderly with Regularly Contacted Network Member		Mean No. of Network Members[a]	
	Men	Women	Men	Women
Network member				
Spouse	64.4	27.4	...	...
Children	65.2	66.0	2.3	2.2
Close relatives	55.3	55.9	3.1	2.9
Total contact with any network	91.6	87.7	4.3	3.8
Cohort size (unweighted no.)	1169	1643	...	...

Note. The family network included the spouse, children, and/or close relatives other than children (including siblings) with whom the respondent had regular face-to-face contact. Data are weighted. EPESE = Established Populations for Epidemiologic Studies of the Elderly.
[a]Means for persons with specified type of network member.

TABLE 3—Adjusted Odds Ratios of Nursing Home Entry, by Family Network Composition and Size, New Haven EPESE, 1982 through 1985

Network Characteristics	Mean (n = 1136)		Women (n = 1603)	
	OR	95% CI	OR	95% CI
Composition				
No spouse/spouse	3.2	1.6, 6.3	2.1	0.9, 4.8
No children/children	1.1	0.4, 3.2	0.7	0.4, 1.4
No relatives/relatives	1.5	0.8, 2.9	0.9	0.6, 1.5
Size				
2 vs 3+	1.8	0.8, 4.3	1.4	0.7, 2.6
1 vs 2	1.2	0.5, 3.0	1.0	0.5, 2.1
0 vs 1	0.8	0.1, 6.4	2.9	1.5, 5.1

Note. The family network included the spouse, children, and/or close relatives other than children (including siblings) with whom the respondent had regular face-to-face contact. Odds ratios (ORs) are adjusted for age, race, housing strata, reported limitations in activities of daily living, and cognitive performance. Data are weighted. CI = confidence interval; EPESE = Established Populations for Epidemiologic Studies of the Elderly.

member of the family network if he or she was reported as a member of the household in 1982.

In 1982, 91.6% of older men and 87.7% of older women in New Haven had regular contact with at least one family member (i.e., spouse, child, or other close relative including siblings). For persons with a family network, the mean family size was 4.3 and 3.8 persons for men and women, respectively. Although overall differences between older men and women are slight with respect to their family networks, the differential in the proportion of men and women with a living spouse is noteworthy.

Results from logistic regression models are presented in Table 3 for men and women. Effects of both the composition and size of the family network on the risk of nursing home entry are examined. The outcome of interest is whether or not a respondent had at least one nursing home admission during the study period. Models control for age, race, housing strata, limitations in activities of daily living, and cognitive performance. Confidence intervals (CIs) for relative risks are based on standard errors adjusted to reflect the sampling design of the study.

For men, the absence of regular contact with a spouse was highly associated with nursing home entry (odds ratio = 3.2, 95% CI = 1.6, 6.3). At the same time, none of the variables indicating the size of the network predicted nursing home entry. Thus, no matter how large the network of relatives, the spouse was most important in reducing the risk of nursing home entry for men.

In contrast, for women, the composition of the network was not particularly important, but the size of the network was predictive of admission. Specifically, the relative risk of entry for elderly women who had no contact with relatives compared with those who had contact with only one relative was 2.9 (95% CI = 1.5, 5.1). This finding suggests that having at least one regularly contacted family member, regardless of his or her relation, reduces the risk of entry for women.

Discussion

In this study, we used prospective community-based data on the elderly population of New Haven to examine the relationship between family networks and the risk of nursing home entry. Our analyses support the hypothesis that, for older persons, the presence of a family network is associated with a reduced risk of nursing home entry, even after controlling for important confounders.

In addition, findings suggest the roles of kin in preventing institutionalization may be distinct for older men and women. For men, network members do not appear to be interchangeable. No matter how large the family, the absence of a spouse nearly triples the risk of nursing home entry, with age, race, functional disability, and mental status controlled. In contrast, the risk of nursing home entry for women appears to be reduced in the presence of at least one family member, regardless of the type of relation.

A major strength of this analysis is that we used a random sample of community-dwelling elderly persons, followed prospectively over time. In addition, we had rich detail on the social network and functional ability of the cohort of elderly persons. This detail enabled us to explore the role of extended family members in nursing home entry.

Nevertheless, our analysis is limited in several ways. We were unable to examine characteristics of network members (i.e., sex, marital status), nor could we identify the relationship of relatives (other than children) to respondents. We also did not take into consideration possible changes in the network over time due to mortality. Further, sample size constraints prohibited any distinction between recuperative and custodial stays.

Despite these limitations, this study points to the importance of regular contact with family members in reducing the risk of nursing home entry for older persons. Health professionals interested in targeting programs to persons at high risk of institutionalization should consider not only functional and mental performance of older persons, but also the extent to which the family is able and willing to provide support in the community. Widowed, divorced, and never-married men are at particularly high risk, as are women without at least one regularly contacted family member. □

Acknowledgments

Funding for this research was provided by Commonwealth Fund Commission on Elderly People Living Alone grant 11186 and also by National Institute on Aging grants N01-AG-0-2105 and T32-AG00153-04.

Helpful comments were made by Daniel Foley on an earlier draft of this paper.

References

1. Liu K, Manton KG. The characteristics and utilization patterns of an admissions cohort of nursing home patients. *Gerontologist.* 1983;23:92–98.
2. Greene VL, Ondrich JI. Risk factors for nursing home admission and exits: a discrete-time hazard approach. *J Gerontol.* 1990;45:S250–S258.
3. Vincent L, Wiley J, Carrington R. The risk of institutionalization before death. *Gerontologist.* 1979;19:361–367.
4. McCoy J, Edwards B. Contextual and sociodemographic antecedents of institutionalization among aged welfare recipients. *Med Care.* 1981;29:907–921.
5. Branch L, Jette A. A prospective study of long-term care institutionalization among the aged. *Am J Public Health.* 1982;80:315–322.
6. Shapiro E, Tate R. Predictors of long term care facility use among the elderly. *Can J Aging.* 1985;4:11–19.
7. Cohen M, Tell E, Wallack S. Client-related risk factors of nursing home entry among elderly adults. *J Gerontol.* 1986;41:S785–S792.
8. Kemper P, Murtaugh C. Lifetime use of nursing home care. *N Engl J Med.* 1991;324:595–600.
9. Palmore E. Total chance of institutionalization among the elderly. *Gerontologist.* 1976;16:504–507.
10. Liu K, Coughlin T, McBride T. Predicting nursing home admission and length of stay: a duration analysis. *Med Care.* 1991;29:125–141.
11. Newman S, Struyk R. Overwhelming odds: caregiving and risk of institutionalization. *J Gerontol.* 1990;45:S173–S183.
12. Foley DJ, Ostfeld AM, Branch LG, Wallace RB, McGloin J, Cornoni-Huntley JC. The risk of nursing home admission in three communities. *J Health Aging.* 1992;4:155–173.
13. Morris JN, Sherwood S, Gutkin CE. Inst-Risk II: an approach to forecasting relative risk of future institutional placement. *Health Serv Res.* 1988;23:511–536.
14. Wolinsky FD, Callahan CM, Fitzgerald JF, Johnson RJ. The risk of nursing home placement and subsequent death among older adults. *J Gerontol.* 1992;47:S175–S182.
15. Steinbech U. Social networks, institutionalization, and mortality among elderly people in the U.S. *J Gerontol.* 1992;47:S183–S190.
16. Townsend P. *The Family Life of Older People.* London, England: Routledge and Kegan Paul, Ltd; 1957.
17. Berkman LF, Berkman C, Kasl S, et al. Depressive symptoms in relation to physical health and functioning in the elderly. *Am J Epidemiol.* 1986;124:372–388.
18. Seeman T, Berkman L. Structural characteristics of social support networks and their relationship with social support in the elderly: who provides support. *Soc Sci Med.* 1988;26:737–749.

The Teaching of Occupational Health in US Medical Schools: Little Improvement in 9 Years

Jay M. Burstein, MD, MPH, and Barry S. Levy, MD, MPH

ABSTRACT

A questionnaire survey of the 127 US medical schools was undertaken to assess the present status of occupational health teaching as a follow-up to two prior similar studies. The present study revealed that 78 (68%) of the 115 responding schools specifically taught occupational health during the 1991/92 academic year, in comparison with 50% in the 1977/78 and 66% in the 1982/83 academic years. The median required curriculum time was 6 hours in 1991/92, as compared with 4 hours in both previous surveys. Despite the increasing recognition of occupational health and growth of information in this field of medicine, occupational health teaching to medical students has not progressed proportionately. (*Am J Public Health.* 1994;84:846–849)

Introduction

Occupational illnesses and injuries are significant problems in the United States. In 1990, there were an estimated 6.4 million work-related injuries and 332 000 new cases of occupational illness.[1] According to the National Safety Council, there were 10 500 work-related fatalities in 1990 alone.[2] These figures are considered gross underestimates of the actual numbers of occupational illnesses and injuries,[3,4] which are widespread and affect virtually all organ systems.[5]

However, the health care system in the United States is not able to meet the challenges of occupationally related and environmentally related illness and injury.[6] A significant contributing factor to this deficiency is the shortage of qualified occupational and environmental health physicians.[7] As of 1990, there was an estimated shortage of 3100 to 5500 physicians in occupational and environmental medicine.[8]

Occupational health is generally defined as the recognition, treatment, and prevention of work-related illnesses and injuries, and occupational medicine is typically defined as occupational health in a clinical context. There are several definitions of environmental health and environmental medicine. Environmental health is often defined as the recognition, treatment, and prevention of illnesses and injuries related to the environment, and environmental medicine is usually defined as environmental health in a clinical context. Some physicians consider occupational health as a subset of environmental health, and others consider these fields as distinct entities.

In recent years, there have been efforts to extend occupational medicine to include environmental medicine. For example, the American College of Occupational and Environmental Medicine and the Association of Occupational and Environmental Clinics are two organizations that focus on both occupational and environmental medicine. In this survey, we solicited responses about the teaching of occupational health and occupational medicine—and, to a limited extent, environmental health and environmental medicine—without defining these terms.

There has, however, been increased interest in occupational and environmental health. The Institute of Medicine has convened several committees to increase the dissemination of information and teaching of occupational medicine to medical students, residents, and physicians.[3,8,9] Several federally funded programs have been initiated to increase and expand occupational teaching and expertise. And the Agency for Toxic Substances and Disease Registry has supported the development of training materials and research fellowships in environmental medicine.[10]

Questionnaire surveys on the teaching of occupational health in US medical schools were performed during the 1977/78 and 1982/83 academic years.[11,12] These prior surveys revealed that occupational health was specifically taught to medical students in only 50% (1977/78) and 66% (1982/83) of medical schools. Occupational health was part of the required core curriculum in 30% (1977/78) and 54% (1982/83) of the schools. The required median time devoted to occupational health was only 4 hours in both surveys.

Physicians can make a considerable impact on the recognition, treatment, and prevention of work-related illness and injury if they are properly trained as students.[6,13–15] The present survey was undertaken to assess how much progress, if any, has been made in recent years in

Jay M. Burstein is with the Occupational Health Program, University of Massachusetts Medical Center, Worcester. At the time of the study, Barry S. Levy was with the Program for Environment and Health, Management Sciences for Health, Boston, Mass.; he is currently with the International Physicians for the Prevention of Nuclear War, Cambridge, and the Tufts University School of Medicine, Boston.

Requests for reprints should be sent to Jay M. Burstein, MD, MPH, Occupational Health Program, University of Massachusetts Medical Center, 55 Lake Ave N, Worcester, MA 01655.

This paper was accepted August 10, 1993.

teaching occupational health to medical students and to identify future needs and directions.

Methods

In September 1991, a questionnaire on the teaching of occupational health during the 1991/92 academic year was sent to all 127 US medical schools. The questionnaire, endorsed by the Association of Teachers of Preventive Medicine, was sent to chairs of departments of preventive and/or community medicine. For the 35 schools without a preventive and/or community medicine department, the questionnaire was sent to chairs of family medicine or, if there was no such department, internal medicine. A cover letter accompanied the survey, requesting that it be forwarded to the person most responsible for teaching occupational health to medical students. Two additional mailings and a telephone call were made to nonrespondents.

The survey elicited the same information as the two earlier surveys, as well as some additional information: name, title, and department of the person who coordinates the teaching of occupational health; the number of faculty members teaching occupational health; whether occupational health was specifically taught in 1991/1992; curriculum content and format; the number of years occupational health has been taught; curriculum requirements, including time devoted to occupational health; elective courses in occupational and/or environmental medicine; and any planned changes for the 1992/93 academic year.

Results

One hundred fifteen (91%) of the 127 medical schools in the United States responded to the survey. Occupational health was specifically taught to medical students in 78 (68%) of responding schools and was a required course or part of a required course in 69 (60%) schools. The median curriculum time devoted to occupational health was 6 hours (mean = 6.7) (Table 1), and occupational health had been specifically taught in these schools for a median of 8 years.

The following results apply only to the 78 schools specifically teaching occupational health, unless otherwise noted.

Coordinator/Director

Information regarding the title and professional qualifications of the coordinator/director of occupational health teaching was surveyed. A professor directed occupational health teaching at 26% of the schools; an associate professor, at 26%; an assistant professor, at 19%; and a "director," at 19%. Other responses included "chief" (3 schools), "head" (1 school), and "senior lecturer" (1 school). The coordinator/director was from a department of family/preventive/community medicine at most schools (65%). Other departments included internal medicine (17%), occupational/environmental health (10%), and epidemiology/public health (8%). At 54 (69%) schools, the coordinator/director was formally trained in occupational health. At 43 (55%) of these 54 schools, the coordinator/director was board certified in occupational health. The median number of faculty members involved in at least some teaching of occupational health was four.

TABLE 1—Occupational Health Teaching in US Medical Schools

	1977/78 (n = 92)	1982/83 (n = 111)	1991/92 (n = 115)
Specifically taught, %	50	66	68*
Part of required curriculum, %	30	54	60*
Time, h			
Median	4	4	6
Mean	5.6	. . .	6.7
Median number of faculty members involved in occupational health	4	4	4

*$P > .05$ (comparing 1982/83 with 1991/92).

TABLE 2—Occupational Health Curriculum Content

	1977/78 (n = 46), %	1982/83 (n = 73), %	1991/92 (n = 78), %
Information on particular work-related diseases	88	90	95
Information on government agencies and programs	67	79	74
Information on health care delivery and preventive medicine in the workplace	67	82	73
Workplace visits	66	60	36[a]
Integration of occupational health into clinical curriculum	38[b]	49[c]	50
Occupational history taking	. . .	. . .	86
Legal aspects of occupational health	. . .	. . .	58
Occupational musculoskeletal disorders	. . .	. . .	58
Occupational respiratory disorders	. . .	. . .	78
Occupational skin disorders	. . .	. . .	50
Occupational neurological disorders	. . .	. . .	58
Occupational reproductive hazards	. . .	. . .	58

[a]n = 77 for this item.
[b]n = 92 for this item.
[c]n = 111 for this item.

Content of Courses or Clerkships

Almost all schools (95%) that specifically taught occupational health included specific information on particular work-related diseases such as occupational musculoskeletal disorders and occupational respiratory disorders in courses or clerkships (Table 2). Of the 78 schools specifically teaching occupational health, 28 (36%) offered visits to factories, clinical laboratories, or other work sites. Of these 28, 10 required written reports of visits by students. During visits, students spoke to health care professionals (17 schools), employers/managers (14 schools), workers (12 schools), and safety or industrial hygiene professionals (20 schools). Occupational health/medicine was integrated into the clinical curriculum in 39 (50%) of the 78 schools, usually during the family practice clerkship and

TABLE 3—Required Courses/Clerkships Incorporating Occupational Health Teaching among Medical Schools in Sample

Course/Clerkship	No.
Preventive medicine	17
Family practice	16
Introduction to clinical medicine	14
Introduction to occupational health	11
Community and preventive medicine	10
Epidemiology/public health	9
Community medicine	8
Physical diagnosis	4
Environmental health	4
Physiology	3
Toxicology	3
Pulmonary medicine	2
Social and behavioral medicine	2
Other	2

occasionally during the internal medicine clerkship.

Elective Curriculum

Sixty-one (78%) of the 78 schools specifically teaching occupational health offered elective courses and/or clerkships in occupational health or occupational medicine. Eight of these 61 schools had no required curricular time offered. Most of the electives were offered in the clinical years. Several were a combination of clinical and research experiences. Several schools offered occupational health as an elective segment of the required family or community clerkship. Other electives included nuclear health hazards, seminars in epidemiology and public health, agricultural medicine, and opportunities in occupational epidemiology. Elective courses in environmental health or environmental medicine were offered in 42 (54%) of the schools. Most of these electives were "environmental health" courses, but several schools combined environmental health with occupational health in an elective course or clerkship.

Research and Summer Experiences for Students

Forty-five (58%) of the 78 schools offered research projects or summer experiences. The median number of students participating in these projects was one (mean = 2.0).

Graduate and Continuing Medical Education

Faculty at 45 (58%) of the 78 schools specifically taught occupational health or occupational medicine to interns or residents. Continuing medical education programs were offered by 43 (55%) of these medical schools, usually as an annual occupational health update course, a seminar, or grand rounds.

Other Aspects

Library services for occupational health were, on average, rated between "fair and good." Occupational health teaching resources and teaching aids were, on average, rated "fair" by the respondents.

Sixty (77%) responding schools operated or were affiliated with an occupational medicine clinic.

Thirty-eight (49%) of the 78 schools specifically teaching occupational health and 8 (22%) of the 37 schools not specifically teaching occupational health planned changes in the 1992/93 academic year, primarily to expand programs.

Discussion

In the past 9 years, there has been almost no change in the percentage of schools specifically teaching occupational health to medical students, although there has been a minor increase in schools requiring occupational health as part of the curriculum (Table 1). In addition, there has been an increase of 2 hours to bring the median number of hours of required occupational health teaching to medical students to 6. These findings are discouraging, given the 9 years since the last survey.

A number of reasons might explain the minimal increases in occupational health education in medical schools during this period. Two possible factors are the lack of increase in faculty qualified to teach occupational health and the paucity of occupational health taught in courses other than preventive medicine and family practice.

The median number of faculty members per school who are teaching occupational health has remained unchanged since 1977/78. More occupational health faculty would probably lead to more occupational health in medical school curricula. Nevertheless, the survey revealed that occupational health was minimally represented in "traditional" medical school courses such as physiology, physical diagnosis, and pathology (Table 3). Only four schools taught occupational health during a physical diagnosis course, an important avenue for introducing occupational and environmental history-taking skills. Occupational health was taught during physiology and toxicology courses at only three schools. Realizing that the present number of faculty teaching occupational health has remained small and that occupational health teaching has a minor presence in the basic science curriculum, we were not surprised by these results.

Improvements can be made if schools exert a sincere commitment to occupational health teaching through empowering existing faculty in this field and providing more opportunities for teaching occupational health. This could occur through a greater number of lectures by faculty currently teaching occupational health. Furthermore, faculty teaching occupational health could educate other existing faculty not currently qualified to teach this subject. However, it will be difficult to substantially recruit future faculty qualified to teach occupational health without necessary funding. Since occupational medicine is not able to generate the same level of income as hospital-based specialties, such as surgery and internal medicine, it is vital that the federal government contribute to recruitment and faculty development in this field. In addition, funding for research activities in occupational health should be increased. Through increasing research activity, more academic occupational health physicians may be recruited, which may, in turn, provide greater opportunity for occupational health faculty to increase teaching within medical schools.

Other potential avenues exist for increasing the occupational health curriculum in medical schools. This could be accomplished through incorporating more occupational health information into the standard preclinical and clinical curricula. For example, information on occupational lung disorders can be taught during pathology or pulmonary medicine courses; occupational musculoskeletal disorders can be taught as a component of rheumatology and orthopedic surgery; toxicology concerning work-related exposures can be incorporated into pharmacology and/or toxicology courses; occupational history taking can be included as a component of physical diagnosis or a patient interviewing course; and occupational reproductive disorders can be taught in endocrinology or obstetrics and gynecology. During

these lectures, handouts and/or case studies could be used. In addition, occupational medicine could become a requirement during a mandatory family or community medicine rotation, as is the case in the curricula of 25 existing medical schools. This concept—incorporation of occupational health teaching into existing curricula—was recommended on several survey questionnaires that we received in response to questions on proposed changes in the teaching of occupational health.

In addition, work-site visits, summer experiences, and research activities provide excellent opportunities for occupational health education. However, in comparison with 1982/83, the present survey revealed a minor decrease in available research projects or summer experiences and a significant decrease in work-site visits. These activities potentially provide an experimental introduction to students unfamiliar with occupational/environmental health and an opportunity for interested students to further develop their knowledge and skills in this field. In addition to summer or research activities at the medical school, opportunities that already exist with the National Institute for Occupational Safety and Health or public health departments may also be made available to interested medical students.

The survey findings also suggest that there is room for improvement in both occupational health teaching resources and aids as well as in occupational health library services. Many of these educational tools are readily obtainable and will be especially helpful for schools that lack faculty with occupational and environmental medicine expertise. Case studies and teaching materials on occupational medicine could be circulated during classroom sessions. In addition, informal journal clubs, discussion groups, or seminars could be established for medical students. This may stimulate the interest of medical students and the commitment of faculty members. Furthermore, occupational health teaching could be included in continuing medical education programs, including grand rounds and didactic lectures in emergency medicine, internal medicine, and family practice departments.

As noted in the results, approximately one half of medical schools are currently teaching environmental medicine. Occupational and environmental medicine educators have recently expressed the need for occupational physicians to evaluate and treat environmental illness, because there are currently no specialists trained in this area of medicine.[16] As more experience in teaching this newly defined area is gained, we encourage the sharing of information among educators.

As more information on occupationally related and environmentally related illness and injury is collected and shared, the need for improving the recognition, treatment, and prevention of occupational illness continues to be evident. This can be better achieved through proper training and the development of needed skills. However, as evidenced by the survey, occupational health teaching in US medical schools continues to be minimal. Recent efforts to expand occupational and environmental health education in medical schools, such as those of the Institute of Medicine and other agencies, may soon reap benefits. Occupational health education in United States medical schools continues to be neglected. A greater commitment is needed to this important subject. □

Acknowledgments

This research was supported in part by a grant from the Association of Teachers of Preventive Medicine.

We acknowledge the assistance and helpful suggestions of Dr Kathleen Rest. We also acknowledge Ms Donna Toney for her work on the manuscript.

References

1. *Bureau of Labor Statistics in the U.S. by Industry, 1990.* Washington, DC: US Dept of Labor; April 1992. Bulletin 2399.
2. *Accident Facts—1991 Edition.* Chicago, Ill: National Safety Council; 1992.
3. Institute of Medicine. *Role of the Primary Care Physician in Occupational and Environmental Medicine.* Washington, DC: National Academy Press; 1988.
4. *Report on Occupational Illness Data Collection: Fragmented, Unreliable, and 70 Years Behind Communicable Disease Surveillance.* Washington, DC: Subcommittee of the Committee on Governmental Operations; 1984.
5. *Handbook of Labor Statistics.* Washington, DC: US Dept of Labor; 1990. Bulletin 2340.
6. Cullen MR, Cherniack MG, Rosenstock L. Occupational medicine. *N Engl J Med.* 1990;322:594–601, 675–683.
7. Rosenstock L, Rest KM, Benson JA, et al. Occupational and environmental medicine: meeting the growing need for clinical services. *N Engl J Med.* 1991;325:924–927.
8. Castorina J, Rosenstock L. Physician shortage in occupational and environmental medicine. *Ann Intern Med.* 1990;113:983–986.
9. Institute of Medicine. *Addressing the Physician Shortage in Occupational and Environmental Medicine.* Washington, DC: National Academy Press; 1991.
10. Institute of Medicine. *Meeting Physicians' Needs for Medical Information on Occupations and Environments.* Washington, DC: National Academy Press; 1991.
11. Levy BS. The teaching of occupational health in American medical schools. *J Med Educ.* 1980;55:18–22.
12. Levy BS. The teaching of occupational health in United States medical schools: five-year follow-up of an initial survey. *Am J Public Health.* 1985;75:79–81.
13. Rosenstock L, Cullen MR. *Clinical Occupational Medicine.* Philadelphia, Pa: WB Saunders Co; 1986.
14. Levy BS, Wegman DH, eds. *Occupational Health: Recognizing and Preventing Work-Related Disease.* 2nd ed. Boston, Mass: Little Brown & Co Inc; 1988.
15. LaDou J. *Occupational Medicine.* Norwalk, Conn: Appleton & Lange; 1990.
16. Cullen MR, Figueroa ST. Incorporating environmental medicine into training programs in occupational medicine. *J Occup Med.* 1990;32:1104–1107.

International Infant Mortality Rates: Bias from Reporting Differences

Embry M. Howell, PhD, and Béatrice Blondel, PhD

ABSTRACT

International infant mortality statistics have caused concern in the United States, since the US ranking relative to other developed countries has declined since World War II. This paper suggests that there may be international differences in reporting of very-low-birthweight infants and perinatal deaths and that such reporting differences bias comparisons of national perinatal and infant mortality rates. Efforts must be made to adopt standard conventions for the inclusion of small, early infants and fetal deaths in rate calculations. (*Am J Public Health.* 1994;84:850–852)

Introduction

Because of their ready availability and recognized importance as health and social indicators, infant mortality rates are routinely used for comparing countries. These statistics have become a source of concern in the United States, because it appears that the degree of decline in US rates has been lower than in other developed countries since World War II.[1]

The usefulness of crude infant mortality rates in international comparisons is questionable because of differences in the registration of births and deaths among very-low-birthweight (<1500 g) infants. In most European Community countries, registration is limited to births and fetal deaths above 27 weeks of gestation.[2] In Norway, registration is at 16 weeks and above,[3] and, in other places such as Osaka Province, Japan, and parts of the United States, all products of conception may be registered even though such registration is certainly incomplete for early fetal deaths. Common practices may also vary at local levels.[4–7] These differences in laws and customs regarding registering fetal deaths and live births have an impact on reported mortality rates.[8,9]

Methods

The purpose of this paper is to present data on birthweight distributions for births and fetal deaths from a wide range of countries and to show how registration and reporting bias may affect mortality rates. We used national samples of births in the United States (1980) and France (1981); Michigan state vital statistics (1984); a birth registry from Lorraine, France (1984); and vital statistics from several countries or regions that were included in the International Collaborative Effort on Infant Mortality sponsored by the US National Center for Health Statistics.[10] The International Collaborative Effort countries and regions were England and Wales; Hungary; Osaka Province, Japan; Norway; Sweden; and several states in the United States: California, Georgia, Michigan, Minnesota, Missouri, North Carolina, Utah, and Wisconsin. To analyze underreporting of very small infants and the possible misclassification of fetal deaths, we also developed hypothetical data to illustrate how reporting may bias international comparisons of infant mortality rates.

Results

Table 1 provides sample survey data, vital statistics, and birth registry data from France and the United States. While French infants were smaller, on average, there were fewer reported very-low-birthweight births proportionately in France. Because mortality rates are extremely high in the very-low-birthweight category (more than 400 per 1000), any underreporting in this category will bias mortality rates. Table 1 also shows that, among fetal and neonatal deaths, the proportion of fetal deaths was substantially higher in Lorraine than in Michigan.

Other data show that in 1985 the United States, Norway, and Osaka Province, Japan, reported between 0.2% and 0.4% of births and fetal deaths as less than 500 g (top panel of Table 2). The proportion of very-low-birthweight infants and fetal deaths ranged from 0.7% in Sweden to 1.5% in Osaka, Japan. Many of these differences are due to differences in reporting requirements for fetal deaths. After excluding all fetal deaths under 28 weeks' gestation (bottom panel of Table 2), we found similar very-low-birthweight rates in Osaka, Norway, and Sweden, while the rates shown before exclusions are quite different.

The proportion of perinatal deaths reported as fetal deaths ranged from 33% in Hungary to 68% in France (Table 3), suggesting large differences across countries in categorizing live births and fetal deaths. These classification differences could bias international infant mortality rate comparisons, since fetal deaths are

Embry M. Howell is with Mathematica Policy Research, Washington, DC. Béatrice Blondel is with the Epidemiology Research Unit on Women and Children's Health, INSERM, Villejuif, Cedex, France.

Requests for reprints should be sent to Embry M. Howell, PhD, Mathematica Policy Research, 600 Maryland Ave, SW, Suite 550, Washington, DC 20024.

This paper was accepted August 3, 1993.

TABLE 1—Low-Birthweight Rates, Mean Birthweight, and Perinatal Mortality Rates: United States and France

	Natality Surveys[a]		Regional Data[b]	
	United States, 1980	France, 1981	Michigan, 1984	Lorraine, 1984
Mean birthweight, g	3365	3322	3274	3250
Very low birthweight, % <1500 g	0.8	0.3	1.3	1.0
Moderately low birthweight, %1500–2499 g	5.0	4.8	5.6	5.5
Perinatal mortality rate,[c] deliveries <1500 g	...	...	428.9	470.1
Perinatal mortality rate,[c] deliveries 1500–2499 g	...	...	51.2	59.2
Perinatal deaths that were fetal deaths, %	...	...	43	61

Source. Raw data were obtained from data tapes of the US Natality Survey, 1980; French Natality Survey, 1981; Michigan Vital Statistics; and Lorraine Birth Registry.
[a]Singleton live births and fetal deaths of 28 weeks' gestation or more.
[b]All live births and fetal deaths of 500 g or more.
[c]Infant deaths in the first 28 days of life plus fetal deaths divided by total births and fetal deaths, multiplied by 1000.

TABLE 2—Very-Low-Birthweight Rates (%) for Selected Countries, 1985

Birthweight, g	England and Wales	Hungary	Japan (Osaka, 1984)	Norway	Sweden	United States
All reported live births and fetal deaths, regardless of gestational age						
<500	0.0	0.0	0.4	0.3	0.0	0.2
500–999	0.3	0.6	0.6	0.4	0.2	0.6
1000–1499	0.7	1.1	0.4	0.5	0.5	0.7
Total <1500	1.0	1.7	1.5	1.2	0.7	1.5
All reported live births and fetal deaths of 28 weeks' gestation or more						
<500	0.0	0.0	0.0	0.0	0.0	0.2
500–999	0.3	0.5	0.2	0.2	0.2	0.5
1000–1499	0.7	1.1	0.4	0.5	0.5	0.6
Total <1500	1.0	1.7	0.6	0.7	0.7	1.3

Source. Data are from International Collaborative Effort on Perinatal and Infant Mortality.[10]

not included in routinely reported infant mortality rates.

Hypothetical data show how underreporting could bias perinatal mortality rates. In a hypothetical country, the actual very-low-birthweight proportion is 1.0%, but the reported percentage is only 0.5. Assume a very-low-birthweight perinatal mortality rate of 470 per 1000 and 100 000 reported live births and fetal deaths. There are actually 470 very-low-birthweight perinatal deaths, but only 235 are reported. The reported perinatal mortality rate is 15 per 1000, but correcting for underreported deaths would raise this rate substantially to 17.3 per 1000.

Another hypothetical example illustrates the impact of differences in the classification of fetal deaths. Two hypothetical countries each have 55 000 live births and fetal deaths, 1000 perinatal deaths, and identical perinatal mortality rates of 18.2 per 1000. However, in the first country 40% of perinatal deaths are classified as fetal deaths, while the proportion is 60% in the second country. Assuming identical postperinatal mortality, the infant mortality rates are very different in the two countries: 14.0 per 1000 in the first country and only 10.4 per 1000 in the second country. This is entirely due to differences in the proportion of fetal deaths among perinatal deaths.

TABLE 3—Proportion of Fetal Deaths among Perinatal Deaths: Selected Countries, 1985

Country	Fetal Deaths, %
England and Wales	56
France	68
Hungary	33
Japan	67
Norway	58
Sweden	54
United States	43

Note. Fetal deaths of 28 weeks' gestation or more and infant deaths in the first 7 days of life are included.
Source. Data are from the International Collaborative Effort on Perinatal and Infant Mortality.[10]

Discussion

Differences in the completeness of reporting of low-birthweight births or fetal deaths and the classification of fetal deaths may substantially bias international comparisons. Such bias was suggested by Lee et al.[11] when they found that the poor ranking of the United States was mainly explained by a high rate of very-low-birthweight babies.

Numerous researchers have stressed the effects of variations in registration limits and the importance of standard rules for international comparisons.[12,13] The World Health Organization recommends that national statistics include all fetuses and infants weighing at least 500 g. In all countries, such recommendations should be used and guidelines developed for health care professionals emphasizing the importance of complete and accurate reporting of vital events. Until reporting becomes more uniform across countries, it is desirable to exclude the smallest birthweight births and to combine infant and fetal deaths before making international comparisons. □

Acknowledgments

This work was funded, in part, by grant HS06621 from the Agency for Health Care Policy and Research, Rockville, Md.

This paper was presented at the 119th Annual Meeting of the American Public Health Association, October 1991, Atlanta, Ga.

Francis Notzon, PhD, of the US National Center for Health Statistics provided access to data on birthweight and infant and fetal

mortality for selected countries and helpful comments on this paper. Paul Vert, MD, provided access to data for Lorraine, France.

References

1. *Health United States 1991.* Hyattsville, Md: US Dept of Health and Human Services; 1992.
2. Kaminski M, Bouvier-Colle MH, Blondel B. *La mortalité des jeunes dans les pays de la communauté européénne.* Paris, France: Doin; 1985.
3. *Births in the Nordic Countries: Registration of the Outcome of Pregnancy: 1978–1983.* Reykjavik, Iceland: Nordisk Medicinal-Statistisk Kommitté; 1987.
4. Doornbos JPR, Nordbeck HJ, Treffers PE. The reliability of perinatal statistics in the Netherlands. *Am J Obstet Gynecol.* 1987; 156:1183–1187.
5. Keirse MJNC. Perinatal mortality rates do not contain what they purport to contain. *Lancet.* 1984;1:1166–1169.
6. Dunn PM, Lashford AM. Limit of fetal viability. *Lancet.* 1984;1:1079–1080.
7. Fenton AC, Field DJ, Mason E, Clarke M. Attitudes to viability of preterm infants and their effect on figures for perinatal mortality. *BMJ.* 1990;300:434–436.
8. Blondel B, Grandjéan BH, Kaminski M, et al. Criteria for registering births. *Lancet.* 1991;337:981.
9. Scott MJ, Ritchie JWK, McClure BG, McReid M, Halliday HL. Perinatal death recording: time for a change? *BMJ.* 1987; 282:707–710.
10. *Proceedings of the International Collaborative Effort on Perinatal and Infant Mortality.* Hyattsville, Md: US Dept of Health and Human Services; 1992; 3.
11. Lee K-S, Paneth N, Gartner LM, Pearlman M. The very low-birthweight rate: principal predictor of neonatal mortality in industrialized populations. *J Pediatr.* 1980;87:758–764.
12. Kline J, Stein Z, Susser M. *Conception to Birth. Epidemiology of Prenatal Development.* New York, NY: Oxford University Press; 1989.
13. Working Group on the Very Low Birthweight Infant. European community collaborative study of outcome of pregnancy between 22 and 28 weeks gestation. *Lancet.* 1990;336:782–784.

Urinary Biological Monitoring Markers of Anticancer Drug Exposure in Oncology Nurses

Mary Anne Newman, PhD, MS, Barbara G. Valanis, DrPH, Rita S. Schoeny, PhD, and Shane Que Hee, PhD

ABSTRACT

People handling anticancer drugs or their wastes may absorb these potent genotoxic agents. The aim of this study was to determine the utility of some general urinary markers among 24 female oncology nurses handling these drugs in comparison with 25 "unexposed" nurses. The markers were the *Salmonella typhimurium* reverse and forward mutation assays, total thioethers, and D-glucaric acid. The reverse mutation assay was the most specific and sensitive marker for anti-cancer drug exposure. Use of the marker battery was no great advantage as a screening tool relative to use of the reverse mutation assay alone. Better recording of work practices in nurse work logs would have improved interpretation of results. (*Am J Public Health.* 1994;84:852–855)

Introduction

Anticancer drugs target cancers because cell division is rapid.[1–3] These drugs affect other proliferating noncancerous tissues (bone marrow, hair follicles, gastrointestinal–nasopharyngeal and genitourinary tract epithelia, and developing embryos).[1–5] Many antineoplastic drugs are carcinogens and teratogens.[4] Adverse effects include irregular menstrual period, testicular function suppression, hair loss, chromosomal abnormalities, decreased white blood cells, liver damage, and spontaneous abortion, as well as light headedness, dizziness, nausea, headache, cough, and skin/mucous membrane and allergic reactions.[1–5] Settings where many of these drugs are administered or prepared (hospitals, home health agencies, pharmacies, waste handlers, and outpatient settings) need sensitive, selective, noninvasive, and inexpensive screening tests reflecting absorption of many anticancer drugs.

Four noninvasive, general urinary antineoplastic drug markers were assessed relative to oncology nurse work practices as part of ongoing research programs.[6–14] The markers were *Salmonella typhimurium* reverse and forward mutagenicity, total thioethers, and D-glucaric acid. All increased for patients receiving high treatment doses.[11] The reverse mutation assay is the "gold standard" for detecting urinary point and frameshift mutagens.[15] About 60% of oncology nurse investigations[1–4,7–10,15–26] are positive. The forward mutation assay detects DNA large base deletions and insertions, base-pair changes, and frameshifts.[27] Thioethers reflect the glutathione detoxification of electrophiles, putative mutagens.[15,28] D-glucaric acid reflects microsomal mixed function oxidation in the liver and kidney and bladder β-glucuronidase activity.[15,29]

Mary Anne Newman is with Health Environments, Inc, Cincinnati, Ohio. Barbara G. Valanis is with the Kaiser-Permanente Center for Health Research, Portland, Ore. Rita S. Schoeny is with the ECAO Division, United States Environmental Protection Agency, Cincinnati, Ohio. Shane Que Hee is with the Department of Environmental Health Sciences, UCLA Center for Occupational and Environmental Health, UCLA School of Public Health, Los Angeles, Calif.

Requests for reprints should be sent to Shane Que Hee, PhD, Department of Environmental Health Sciences, UCLA School of Public Health, 10833 Le Conte Ave, Los Angeles, CA 90024-1772.

This paper was accepted September 30, 1993.

Materials and Methods

Populations and Sampling

Twenty-four "exposed" female oncology nurses and 24 "unexposed" female control nurses were selected from a larger study of health effects[7] associated with antineoplastic drug handling. A physical examination and questionnaires (demographic, general health, smoking, medication use, and perceived exposure data) were administered, and the nurses maintained work logs (frequency, extent, sources, and settings of exposure data).[7] Work practices were observed directly.

Spot (untimed) urine samples were collected in 1983 during physical examinations. The pH-adjusted and filtered urines[12] were frozen at −20°C.

Urinary Markers

Analysis occurred in 1985. Creatinine was measured by Du Pont Automatic Analyser[11] to account for dilution. Thioether and D-glucaric acid concentrations were determined with our optimized methodology.[11] Reverse mutations were measured with *S. typhimurium* strains TA1535 (point mutations) and TA98 (frameshift mutations)[11,12]; forward mutations were quantified with *S. typhimurium* strain TM677 clone 09.[11,12] Assays were with (+) and without (−) rat liver microsomal S9 fraction[12] to indicate the role of metabolic activation.

Statistical Treatments

The creatinine, mutagenicity (reverse and forward +S9 or −S9), thioether, and D-glucaric acid data were stratified in terms of smoking status. Chi-square tests for normality of raw and log-transformed data were used to indicate the appropriate representative values to be calculated.[30] Arithmetic means and standard deviations were computed for data obeying a normal distribution, whereas geometric means and geometric standard deviations were calculated for lognormal distributions. Analysis of variance (ANOVA) techniques allowed significant differences and interactions to be determined[30] relative to anticancer drug exposure and smoking status. Nonparametric tests and regression correlational analyses were used as appropriate.

Results

There was no statistically significant age difference between the oncology nurses (age range = 24 to 54 years, average age = 36 years) and the controls (age range = 23 to 61 years, average age = 40 years). Forty-six percent of the latter and 25% of the former were smokers.

Direct observation revealed the following potential exposure situations.

- During preparation: Powder particles and liquid droplets aerosolize. Also, spills, leaks, and container/syringe breakage occur during transport from or to the pharmacy or during shaking.
- During administration: Syringes leak during transport, priming of intravenous sets, expelling of air, and connection to or removal from the patient. Aerosols form during priming, expelling of air, and connection to or removal from the patient.
- Miscellaneous exposures. Discarded containers contaminate housekeeping workers. Also, improperly cleaned equipment/containers and patient excreta are sources of contamination.

TABLE 1—Reverse Mutagenicity Results for Oncology Nurses and Control Nurses

Group	No.[a]	Smoker	Weight[b] (% SE)[c]	Geometric Mean (% SE)[c] Revertant Colonies per Plate: TA 1535 −S9	TA 1535 +S9[d]	TA 98 −S9
Oncology nurses	18	No	66.1 (6.6)	11.2 (8.2)	8.51 (9.8)	15.4 (7.6)
	6	Yes	103 (3.7)	34.1 (19)	30.0 (19)	46.1 (15)
Control nurses	13	No	81.3 (5.4)	14.1 (7.6)	9.10 (5.9)	13.1 (12)
	11	Yes	77.6 (2.9)	10.2 (5.5)	6.71 (5.6)	13.4 (12)

Note. Detailed results are available from the authors on request.
[a]Totals may vary because of missing data.
[b]Residue dry weight after column chromatography in mg/100 mL urine.[12]
[c]100 (log[geometric mean standard error]/log geometric mean).
[d]Sprague Dawley liver S9 volume after induction with Aroclor 1254 was 20 μL/plate.[12]

Individual Tests' Results

Reverse mutation assay. Of 19 oncology nurses, 3 produced urine mutagenic with strain TA1535 +S9 or −S9 and with TA98 −S9 (Table 1). None of 19 control nurses had urine mutagenic with TA1535 +S9 or −S9; 5 of 11 control group smokers had urines slightly mutagenic with strain TA98 −S9 but not on dilution. The data were lognormally distributed.[30] The sensitivity, specificity, and efficiency proportions, assuming that all oncology nurses absorbed anticancer drugs, were 16%, 100%, and 58%, respectively. The test was very specific but insensitive.

Forward mutation assay. Mutagenicity in the absence of S9 was found for 13 of 15 nonsmokers and all 6 smokers in the oncology nurse group (Table 2). In the control nurse group, 8 of 10 nonsmokers and 8 of 11 smokers had mutagenic urine. The values of sensitivity, specificity, and efficiency, assuming that the oncology nurses absorbed anticancer drugs, were 90%, 23%, and 57%, respectively. The test was sensitive but not very specific.

In +S9 assays for oncology nurses, 2 of 15 nonsmokers and 4 of 6 smokers had mutagenic urine. For the controls, 5 of 10 nonsmokers and 4 of 11 smokers had mutagenic urine. The sensitivity, specificity, and efficiency proportions were 29%, 52%, and 43%, respectively, assuming that the oncology nurses absorbed anticancer drugs.

All data were lognormally distributed.[30] Dose responses on dilution were observed only in S9 absence.

Creatinine. Nonparametric ANOVA on ranks showed no significant differences ($P < .05$) between the averages of the data for the two nurse populations or between smokers and nonsmokers; also, there were no interactions between exposure and smoking status. Five oncology nurses and two controls had dilute[31] urines less than 2.65 mM.

D-Glucaric acid. The nonparametric median test for 19 oncology nurses and 22 control nurses showed no significance ($P < .05$) between the number of nurses above the total group median, regardless of smoking status. Three oncology nurse urines exceeded the highest control value. Concentrations for US females range from 20.2 to 42.0 μmol/g creatinine.[11] Assuming that anticancer drugs were

TABLE 2—Forward Mutagenicity Results with *Salmonella typhimurium* Strain TM 677 for Oncology Nurses and Control Nurses

Group	No.[a]	Smoker	Weight[b] (% SE)[c]	Mean (% SE)[c] −S9 Mutants per 10^8 Survivors	−S9 Viability %	+S9[d] Mutants per 10^8 Survivors	+S9[d] Viability %
Oncology nurses	18	No	65.2 (6.7)	9.03 (5.9)	84.9 (3.2)	6.09 (9.9)	83.6 (1.8)
	6	Yes	106 (4.1)	13.5 (11)	69.9 (3.4)	8.95 (11)	92.8 (1.4)
Control nurses	13	No	102 (4.1)	8.41 (14)	92.0 (2.8)	6.90 (9.9)	85.0 (1.7)
	11	Yes	77.8 (3.0)	9.85 (7.3)	84.1 (1.4)	5.49 (9.7)	95.9 (1.3)

Note. Detailed results are available from the authors on request.
[a]Totals may vary because of missing data.
[b]Residue dry weight after column chromatography in mg/100 mL urine.[12]
[c]100 (log[geometric mean standard error]/log geometric mean).
[d]Sprague Dawley liver S9 volume after induction with Aroclor 1254 was used at 1% (v/v).[12]

absorbed by all oncology nurses, the sensitivity, specificity, and efficiency values of the test relative to the 42 μmol/g creatinine upper reference limit were 33%, 67%, and 50%, respectively.

Thioethers. There was no statistical difference ($P < .05$) between the group means of control nurses and of oncology nurses, regardless of smoking status. No oncology nurse urine exceeded the highest thioether concentration for the controls. The recommended upper limit[28] is 5.9 mmol thiols(−SH)/mol creatinine. Two control nurses exceeded this limit. Assuming that the oncology nurses absorbed anticancer drugs, the sensitivity, specificity, and efficiency proportions relative to the 5.9 mmol −SH/mol creatinine upper reference limit were 0%, 92%, and 46%, respectively. The test was specific but not sensitive.

Marker cross comparison. There were no correlations (Pearson $P < .05$) between any of the markers relative to smoking status, except individually for the three oncology nurses whose urines exhibited reverse mutation. There were also no correlations between marker concentrations and work practices. Smoking alone does not explain the incidence of forward mutations. S9 addition increases the specificity of the forward mutation assay.

One oncology nurse's urine had high glucaric acid and relatively high thioethers (5.68 mmol −SH/mol creatinine) and showed forward and reverse mutation with TA98 before dilution. She also probably absorbed antineoplastic drugs. Thus, only four oncology nurses excreted enough of these drugs to be detected by this battery of markers.

Discussion

This study is the first to report forward mutation assay data for oncology nurses handling anticancer drugs. The assay was sensitive but nonspecific in the absence of S9. Addition of S9 caused greater specificity but markedly decreased sensitivity.

Use of the marker battery was no great advantage as a screening tool relative to the reverse mutation assay alone, since only one extra oncology nurse was judged to have absorbed anticancer drugs.

The mutagenicity data were lognormally distributed for these weakly mutagenic spot urines. Some investigators have assumed that their data are distributed normally.[17,25] A lognormal distribution describes the incidence of micronuclei in human blood peripheral lymphocytes exposed in vitro to adriamycin and cyclophosphamide in urines and exposed to 24-hour urines from patients given high doses of anticancer drugs.[13] At high mutagen concentrations, as found for concentrated 24-hour urines, the data are distributed normally.[12]

The lack of correlation between work practices and detection of urinary mutagenicity was surprising. The absolute sensitivity of the reverse mutation assay is more than adequate to detect mutagens in urine.[1–4,7–12,15–26]

The sensitivity of the urinary mutagenicity marker for these spot urine samples is, however, dependent on the pharmacokinetics of the drugs. Urinary mutagenicity for chemicals of short body half-times (e.g., cyclophosphamide) after intravenous injection will provide information about same-day or recent exposures. Cyclophosphamide was not detected in these urines, since addition of S9 did not result in the expected enhanced mutagenicity for strain TA1535.[12] Conversely, exposures to drugs of long body half-times such as adriamycin[1–3,15] will not be correlated to recent exposures. The whole body half-time after skin exposure is generally much longer than after oral or intravenous dosing. This suggests that 12- or 24-hour urines should be collected instead of spot samples.

Skin absorption is the major route of anticancer drug exposure.[4,5] The rate and efficiency of absorption is a function of skin location, exposure time, skin exposure area, the presence and protective efficiency of gloves and garments, skin blood perfusion, and the thickness of fat layers under the epidermis.[15] Since most of the anticancer drugs are administered in aqueous solution, long body half-times are expected, even for chemicals of short body half-time when injected intravenously, since intact skin is protective against hydrophilic and very polar molecules.[15] Thus, the time and duration of urine sampling may be crucial to the detection of these urinary mutagens as well as of the other urinary markers. This may explain the general insensitivity of the entire battery of markers (apart from the nonspecific forward mutation assay in the absence of S9) and why mutagenicity did not correlate with potential exposure events on the same day. Thioethers are also known to increase some time after anticancer drug exposure.[32–34]

The self-reported information in the nurse work logs was also not detailed enough to enable assessment of the intensity and duration of any direct exposure of the skin. This is not surprising since nurses are very busy and instant recording is difficult. Later recall of spills under such conditions would be even worse. This suggests that experienced observers who know how to account for the effectiveness of protective equipment must be used to estimate exposure extent. This should be done in future studies.

The reverse mutation assay is the most sensitive and specific assay to detect urinary mutagens after antineoplastic drug exposure. More research is necessary to determine how skin exposure correlates to urinary mutagenicity and when and for what duration urine sampling should be performed after dermal exposure. □

References

1. Remers WA, ed. *Antineoplastic Agents.* New York, NY: Wiley, 1984.
2. Calabresi P, Parks RE Jr. Chemotherapy of neoplastic diseases. In: Gilman AG, Goodman LS, Rall TW, Murad F, eds. *Goodman and Gilman's The Pharmacological Basis of Therapeutics.* 7th ed. New York, NY: Macmillan; 1985:1240–1306.
3. Crooke ST, Prestayko AW, eds. *Cancer and Chemotherapy, Volume III, Antineoplastic Agents.* New York, NY: Academic Press; 1981.
4. *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans.* Lyon, France: International Agency for Research on Cancer; 1981;26:37–384, 441.
5. *Work Practice Guidelines for Personnel Dealing with Cytotoxic (Antineoplastic) Drugs.* Washington, DC: Occupational Safety and Health Administration; 1986.
6. Valanis B, Browne M. Use of protection by nurses during occupational handling of antineoplastic drugs. *Natl Intravenous Ther Assoc.* 1985;8:218–222.
7. Valanis BG, Hertzberg V, Shortridge L. Antineoplastic drugs. Handle with care. *Amer Assoc Occup Health Nurses J.* 1987;35: 487–492.
8. Valanis B, Shortridge L. Self-protective practices of nurses handling antineoplastic drugs. *Oncol Nurs Forum.* 1987;14:23–27.
9. Valanis B, Driscoll K, McNeill V. Comparison of antineoplastic handling policies of hospitals with OSHA guidelines: a pilot study. *Am J Public Health.* 1990;80:480–481.
10. Valanis B, McNeil V, Driscoll K. Staff members' compliance with their facility's antineoplastic drug handling policy. *Oncol Nurs Forum.* 1991;18:571–576.
11. Newman MA, Que Hee SS, Schoeny R, Lowry L. Biological monitoring screening of patients provided antineoplastic drugs including adriamycin, cyclophosphamide, 5-fluorouracil, methotrexate, and vincristine. *Cancer Res.* 1990;50:3351–3366.
12. Newman MA, Que Hee SS, Schoeny RS. Mutagenesis assays on urines produced by patients administered adriamycin and cyclophosphamide. *Environ Mol Mutagen.* 1990; 15:189–203.
13. Boucher R, Livingston GK, Que Hee SS. The in-vitro micronucleus bioassay of human peripheral lymphocytes applied to adriamycin, cyclophosphamide, and urines of patients administered anticancer drugs. *Environ Mol Mutagen.* 1993;21:372–382.
14. Igwe OJ, Que Hee SS, Wagner WD. Urinary thioether biological monitoring in the interaction between 1,2-dichloroethane and disulfiram in Sprague-Dawley rats. *Am Ind Hyg Assoc J.* 1988;49:10–16.
15. Que Hee SS. *Biological Monitoring: An Introduction.* New York, NY: Van Nostrand Reinhold; 1993.
16. Falck K, Grohn P, Sorsa M, Vainio H, Heinonen E, Holsti LR. Mutagenicity in urine of nurses handling cytostatic drugs. *Lancet.* 1979;1:1250–1251.
17. Ratcliffe JM. *Occupational Exposure to Cancer Chemotherapeutic Agents in Pharmacists and Nurses.* Washington, DC: National Institute for Occupational Safety and Health, 1983.
18. Jones R, Frank R, Mass T. Safe handling of chemotherapeutic agents: a report from the Mount Sinai Medical Center. *CA.* 1983;33:258–263.
19. Staiano N, Galleli JF, Adamson RH, Thorgeisson SS. Lack of mutagenic activity in urine from hospital pharmacists administering antitumor drugs. *Lancet.* 1981;1:615–616.
20. Cloak M, Connor T, Stevens K, et al. Occupational exposure of nursing personnel to antineoplastic agents. *Oncol Nurs Forum.* 1985;12:33–39.
21. Rogers B, Emmett EA. Handling antineoplastic agents: urine mutagenicity in nurses. *Image.* 1987;19:108–113.
22. Wilson JP, Solimando DA. Antineoplastics: a safety hazard? *Am J Hosp Pharm.* 1981;38:624.
23. Everson RB, Ratcliffe JM, Flack PM, Hoffman DM, Watanabe AS. Detection of low levels of urinary mutagen excretion by chemotherapy workers which was not related to occupational drug exposures. *Cancer Res.* 1985;45:6487–6497.
24. Anderson RW, Puckett WH, Dana WJ, Nguyen TV, Theiss JC, Matney TS. Risk of handling injectable antineoplastic agents. *Am J Hosp Pharm.* 1982;39:1881–1887.
25. Poyen D, De Meo MP, Botta A, Gouvernet J, Dumenil G. Handling of cytostatic drugs and urine mutagenesis. *Int Arch Occup Environ Health.* 1988;61:183–188.
26. Sorsa M, Pyy L, Salomaa S, Nylund L, Yager JW. Biological and environmental monitoring of occupational exposure to cyclophosphamide in industry and hospitals. *Mutagen Res.* 1988;201:165–180.
27. Skopek TR, Liber HL, Krolewski JJ, Thilly WG. Quantitative forward mutation assay in *Salmonella typhimurium* using a 8-azaguanine resistance as a genetic marker. *Proc Natl Acad Sci USA.* 1978;75:410–414.
28. van Doorn R, Leijdekkers CM, Bos RP, Brouns RME, Henderson PT. Detection of human exposures to electrophilic compounds by assay of thioether detoxification products in urine. *Ann Occup Hyg.* 1981;24: 77–92.
29. Notten WRF, Henderson PT. The interaction of chemical compounds with the functional state of the liver. II. Estimation of changes in D-glucaric acid synthesis as a method for diagnosing exposure to xenobiotics. *Int Arch Occup Environ Health.* 1977;38:209–220.
30. Snedecor GW, Cochran WG. *Statistical Methods.* 8th ed. Ames, Iowa: Iowa State University Press; 1989.
31. Lauwerys RR. *Industrial Chemical Exposure: Guidelines for Biological Monitoring.* Davis, Calif: Biomedical Publications; 1983.
32. Jagun O, Ryan M, Wadron HA. Urinary thioether excretion in nurses handling cytotoxic drugs. *Lancet.* 1982;2:443–444. Letter.
33. Burgaz S, Ozdamar YN, Karakaya AE. A signal assay for the detection of genotoxic compounds: application on the urines of cancer patients on chemotherapy and of nurses handling cytotoxic drugs. *Hum Toxicol.* 1988;7:557–560.
34. Sarto F, Trevisan A, Tomanin R, Canova A, Fiorentino M. Chromosomal aberrations, sister chromatid exchanges, and urinary thioethers in nurses handling antineoplastic drugs. *Am J Ind Med.* 1990;18: 689–695.

The Consumption of Well-Done Red Meat and the Risk of Colorectal Cancer

Joshua E. Muscat, MPH, and Ernst L. Wynder, MD

ABSTRACT

Heterocyclic aromatic amines and polycyclic aromatic hydrocarbons are mutagens that are produced in highly cooked meats. A case–control study of 511 patients with colorectal cancer and 500 matched control subjects examined whether consumption of well-done cooked beef is related to the risk of developing large bowel cancer. Approximately 16% of men and women consumed well-done beef, and 50% ate medium-cooked beef. For both sexes, there was no association between consumption of well-done or medium-cooked beef and colorectal cancer. This paper discusses whether questionnaire data accurately reflect dietary intake of heterocyclic aromatic amines and polycyclic aromatic hydrocarbons. (*Am J Public Health.* 1994;84:856–858)

Introduction

Heterocyclic aromatic amines and polycyclic aromatic hydrocarbons are mutagens that are formed during high-temperature cooking of red meat and, in lesser amounts, white meat and fish.[1] Well-done meat contains 10 times the concentration of heterocyclic aromatic amines than rare meat prepared by the same cooking method.[2] These amines are mutagenic in bacterial assays[3] and have produced colon tumors in laboratory animals.[4]

Only a few epidemiological studies have examined the association between cooked red meats and colon or rectal cancer. In a prospective study of 89 000 US nurses, Willett et al.[5] reported no association between cooked red meat that was consumed daily and colon cancer; however, they provided no data. In a case–control study, Lyon and Mahoney[6] found no relationship between consumption of broiled or fried meats and colon cancer. In contrast, a Swedish case–control study of colorectal cancer found an increased risk with total meat intake and a preference for heavily browned meat surface.[7] Schiffman and Felton[8] postulated that the degree of doneness of red meat may be an important indicator of heterocyclic aromatic amine intake, and they provided preliminary results from a case–control study of colorectal cancer (n = 50 cases) that showed a significantly increased risk among subjects who consumed well-done red meat. We examined the association between the consumption of well-done beef and colorectal cancer using data from a large case–control study.

Methods

Five hundred eleven patients with histologically confirmed primary cancer of the large bowel (346 with colon cancer, 165 with rectal cancer) were interviewed between 1989 and 1992 in five hospitals. Control patients included 500 subjects matched randomly with each patient by sex, race, hospital, age (within 5 years), and time of the case interview (within 2 months). Eligible control patients included patients with conditions unrelated to dietary fat and fiber intake, such as joint replacement (9.1%), injuries (8.7%), benign prostatic hyperplasia (7.9%), some musculoskeletal disorders (6.7%), melanoma (4.8%), benign skin diseases (4.0%), ill-defined symptoms (3.6%), renal calculus (3.2%), and hernia (2.8%). The pathology reports of the cases were examined to determine the site and stage of the tumor. More than 90% of both case and control patients who were approached participated in the study.

Data on sociodemographic characteristics, medical history, analgesic use, family history of cancer, smoking, alcohol use, physical activity, vitamin use, and dietary history were collected using a standardized questionnaire administered by trained interviewers. The food frequency section included questions on the major sources of total dietary fat and fiber. For the red meat questions, patients were asked how many times they ate "beef, steaks, roasts or hamburgers" per day, week, or month "during the longest period of your adult life," and whether they ate their beef "well-done, medium or rare."

Relative risks of colorectal cancer associated with well-done beef were calculated using the ratio of discordant case–control pairs for men and women.[9] Chi-square analysis and 95% confidence intervals (CIs) were calculated to test statistical significance.

Results

Table 1 presents basic sociodemographic characteristics of the study subjects. Case and control patients had similar ages and years of education. A higher proportion of female control pa-

The authors are with the Division of Epidemiology at the American Health Foundation, New York, NY.

Requests for reprints should be sent to Joshua E. Muscat, MPH, Division of Epidemiology, American Health Foundation, 320 E 43rd St, New York, NY 10017.

This paper was accepted October 4, 1993.

tients than of female case patients were never married. Male control patients were more likely to have been Jewish and less likley to have been Catholic than male case patients. Approximately 50% of men and women in both groups combined ate medium-cooked beef, and 16% had their beef cooked well-done. The frequency of beef consumption and other potential risk factors were unrelated to the degree of doneness.

Among men, the relative risks for colorectal cancer were 1.15 for consumption of well-done beef (95% CI = 0.6, 2.4) and 1.00 for consumption of medium-cooked beef (Table 2); for rectal cancer only, the relative risk for well-done beef was 2.0 (95% CI = 0.7, 5.9). For women, the relative risks for colorectal cancer were 1.00 for consumption of well-done beef and 0.95 for consumption of medium-cooked beef, with no differences in risk between the colon and the rectum.

Discussion

International correlations between per capita consumption of meat and colorectal cancer rates, as well as epidemiological studies of dietary fat intake, have implicated fat in colon cancer carcinogenesis.[10] Schiffman and Felton[8] suggest that these correlations may actually reflect intake of heterocyclic aromatic amines. If both fat and amine intake are risk factors for large bowel cancer, then isolating the effects due to amine intake may be difficult in epidemiological studies. The fat content of red meat diminishes with the amount of cooking. Well-done beef contains 36% fat by weight, whereas rare or raw beef contains more than 50% fat according to one study.[11] Therefore, the potentially higher risk associated with well-done beef may be counterbalanced by the lower fat content of the meat. On the other hand, low-quality beef with higher fat content is usually cooked more thoroughly than prime cuts and with additional sauces or gravy (which may also contain fat and heterocyclic aromatic amines) to add flavor and texture.

The content of heterocyclic aromatic amines in beef increases substantially with frying time,[12] suggesting that substantial amounts of amines are present only in well-done beef. Schiffman and Felton[8] report that subjects who consumed well-done beef had a 3.5-fold (95% CI = 1.3, 9.6) increased risk of developing colorectal cancer compared with subjects who ate "rare to medium rare" red meat. The study by deVerdier et al.[7] shows that consumption of heavily browned meat increases the risk of colorectal cancer. Lyon and Mahoney[6] found no significant associations between consumption of broiled or fried meat and risk of colon

TABLE 1—Characteristics of 511 Colorectal Case Subjects and 500 Control Subjects

	Men				Women			
	Case (n = 283)		Control (n = 276)		Case (n = 228)		Control (n = 224)	
	No.	%	No.	%	No.	%	No.	%
Age, y								
<45	13	4.6	15	5.4	11	4.8	12	5.4
45–54	21	7.4	23	8.3	34	14.9	31	13.8
55–64	75	26.5	74	26.8	57	25.0	58	25.9
65–74	124	43.8	136	49.3	87	38.2	98	43.8
>74	50	17.7	28	10.1	39	17.1	25	11.2
Education, y								
<12	51	18.0	40	14.5	37	16.2	33	14.7
12	94	33.2	79	28.6	102	44.8	93	41.5
13–15	56	19.8	60	21.7	46	20.2	50	22.3
16	33	11.7	45	16.3	20	8.8	24	10.7
>16	48	17.0	52	18.8	22	9.7	24	10.7
Marital status								
Single	18	6.4	10	3.6	11	4.8	21	9.4*
Married	228	80.6	216	78.3	127	55.7	110	49.1
Divorced	20	7.1	19	6.9	19	8.3	33	14.7
Widowed	3	1.1	5	1.8	6	2.6	4	1.8
Separated	14	4.9	26	9.4	65	28.5	56	25.0
Religion								
Protestant	61	21.6	49	17.8**	51	22.4	57	25.4
Catholic	130	45.9	110	39.9	91	39.9	80	35.7
Jewish	86	30.4	103	37.3	77	33.8	77	34.4
Other/none	6	2.1	14	5.1	9	3.9	10	4.5
Beef preparation[a]								
Rare	90	32.4	82	30.3	90	40.2	83	37.9
Medium	137	49.3	150	55.3	97	43.3	102	46.6
Well-done	51	18.3	39	14.4	37	16.5	34	15.5

[a]Excludes 10 male and 9 female subjects who do not eat red meat; across the columns, bases are 278, 271, 224, and 219.
*$\chi^2_{4df} = 9.14, P < .07$; **$\chi^2_{3df} = 8.7, P < .05$.

TABLE 2—Matched-Pair Analysis of Degree of Red Meat Doneness and Colorectal Cancer

Case Subjects	Control Subjects			Relative Risk	95% Confidence Interval
	Rare	Medium	Well-Done		
Men					
Rare	27	42	13	1.00[a]	. . .
Medium	42	72	19	1.00	. . .
Well-done	15	34	5	1.15	0.6, 2.4
Women					
Rare	27	44	12	1.00[a]	. . .
Medium	42	33	14	0.95	0.6, 1.5
Well-done	12	15	8	1.00	. . .

Note. Double-underlined numbers are the discordant pairs for well-done meat. Single-underlined numbers are the discordant pairs for medium-done meat. Table excludes subjects who did not have a matched pair or whose pair did not eat red meat.
[a]Reference category. An unmatched analysis yielded the same results.

cancer. These discrepant study findings may reflect differences in the method of cooking beef (e.g., frying and broiling vs baking) and in the amount of well-done beef consumed. There are little population-based data on beef cooking habits, although one case–control study of 7860 subjects found that men and women ate meat approximately 22 and 18 times per month, respectively,[13] compared with 13 and 11 times per month, respectively, in this study. The lack of an association in our study may reflect a relatively low intake of highly cooked beef.

Questions on meat consumption frequency tend to have low reliability.[14,15] It seems intuitive that subjects would respond more accurately to questions on beef doneness, but this needs to be verified. In the Swedish colorectal cancer study (n = 559 cases), a significant risk was found for high "temperature when frying" and "heavily browned" meat surface,[7] although these definitions may be especially prone to subjective responses. In the study by Schiffman and Felton,[8] a significant risk was found for well-done red meat compared with both red meat cooked "medium to well" and rare red meat, yet the actual difference in heterocyclic aromatic amine intake between "well-done" and "medium to medium-well" meat may be slight. Further research in this area should consider the opposing risks of a higher amine amount but lower fat content in well-done beef and other meats, and should determine the reliability and validity of questions on meat cooking habits. □

Acknowledgments

The authors thank the following individuals and institutions for their collaborative efforts: Mark Citron, MD, Long Island Jewish Medical Center; Robert Burakoff, MD, Winthrop University Hospital; Christine Johnson, MD, Henry Ford Hospital; Ronald Blum, MD, New York University Medical Center; Alfred Neugut, MD, Columbia University; and Randall Harris, MD, PhD, Ohio State University. We also thank Drs Rashmi Sinha and Mark Schiffman for expert comments, and interviewers Cindy Patterson, Kathy Mehl, Rebecca Moldwin, and Marion Rogers.

References

1. Adamson RH. Mutagens and carcinogens formed during cooking of foods and methods to minimize their formation. In: DeVita V, Hellman S, Rosenberg SA, eds. *Cancer Prevention.* Philadelphia, Pa: JB Lippincott Co; November 1990:1–7.
2. Bjeldanes LF, Morris MM, Felton JS, et al. Effect of meat composition and cooking conditions on mutagen formation in fried ground beef. *J Agriculture Food Chem.* 1983;31:18–21.
3. Overvik E, Gustafsson JA. Cooked-food mutagens: current knowledge of formation and biological significance. *Mutagenesis.* 1990;5:437–446. Review.
4. Takayama S, Nakatsuru Y, Masuda M, et al. Demonstration of carcinogenicity in F344 rats of 2-amino-3-methyl-imidazo[4,5-f]quinoline from broiled sardine, fried beef, and beef extract. *Gann.* 1984;75:467.
5. Willett WC, Stampfer MJ, Colditz GA, Rosner BA, Speizer FE. Relation of meat, fat, and fiber intake to the risk of colon cancer in a prospective study among women. *N Engl J Med.* 1990;323:1664–1672.
6. Lyon JL, Mahoney AW. Fried foods and the risk of colon cancer. *Am J Epidemiol.* 1988;128:1000–1006.
7. de Verdier MG, Hagman U, Peters RK, Steineck G, Overik E. Meat, cooking methods and colorectal cancer: a case-referent study in Stockholm. *Int J Cancer.* 1991;49:520–525.
8. Schiffman MH, Felton JS. Re: "Fried foods and the risk of colon cancer." *Am J Epidemiol.* 1990;131:376–378.
9. Breslow NE, Day NE. *Statistical Methods in Cancer Research.* Vol 1: *The Design and Analysis of Case-Control Studies.* Lyons, France: International Agency for Research on Cancer; 1980.
10. Carroll KK. Dietary fats and cancer. *Am J Clin Nutr.* 1991;53:1064S–1067S.
11. Wynder EL, Reddy B. Diet and cancer of the colon. In: Winick M, ed. *Nutrition and Cancer.* New York, NY: John Wiley & Sons Inc; 1977:55–71.
12. Weisburger JH, Spingarm NE. Mutagens as a function of mode of cooking of meats. In: Miller EC, ed. *Naturally Occurring Carcinogens-Mutagens and Modulators of Carcinogens.* Baltimore, Md: University Park Press; 1979:177–184.
13. Morabia A, Wynder EL. Dietary habits of smokers, people who never smoked, and ex-smokers. *Am J Clin Nutr.* 1990;52:933–937.
14. Nomura AMY, Hankin JH, Rhoads GG. The reproducibility of dietary intake data in a prospective study of gastro-intestinal cancer. *Am J Clin Nutr.* 1976;29:1432–1436.
15. Morabia A, Moore M, Wynder E. Reproducibility of food frequency measurements and inferences from a case-control study. *Epidemiology.* 1990;1:305–310.

Cancer Control Career Development Award for Primary Care Physicians

The American Cancer Society is pleased to announce the 1995 Cancer Control Career Development Award for Primary Care Physicians. This award is intended to develop academic leaders in primary care specialties emphasizing cancer control: family practice, general internal medicine, obstetrics and gynecology, and pediatrics. Through the award, the Society seeks to support individuals in supervised programs that will develop the candidate's clinical and teaching expertise and his or her capacity to perform independent clinical research in cancer control. It is anticipated that physicians trained under these awards will improve cancer control through involvement in primary care practice, education, and research activities related to cancer control.

This is a 2-year award. The stipend for the award is $25 000 for the first year and $30 000 for the second. Candidates for first-year awards may not have an academic rank above that of assistant professor and must not be tenured or be the section head (or equivalent) in their discipline. These awards are intended to support the early development of academic careers that place emphasis on cancer control; physicians with well-established careers and substantial research funding should not apply. In addition, candidates for these awards must be citizens or permanent residents of the United States.

The application deadline is August 15, 1994, for awards to begin July 1, 1995. For brochures, applications, and further information, please contact the office of Virginia Krawiec, MPA, Detection and Treatment Department, American Cancer Society, Inc, 1599 Clifton Rd, NE, Atlanta, GA 30329-4251; tel (404) 329-5734; fax (404) 325-1467.

Outbreak and Sporadic Egg-Associated Cases of *Salmonella enteritidis:* New York's Experience

Dale L. Morse, MD, MS, Guthrie S. Birkhead, MD, MPH, Joanne Guardino, MS, Stanley F. Kondracki, and John J. Guzewich, MPH

ABSTRACT

Since 1985, egg-associated *Salmonella enteritidis* has emerged as a major cause of foodborne disease. New York State has been especially affected, with 47 documented egg-associated *S enteritidis* outbreaks involving 2279 cases and 10 deaths. Individual case reports of salmonella have also increased 56%, and sporadic cases of *S enteritidis* have been shown to be associated with egg consumption. Further educational and regulatory activities are needed to control this continuing public health problem. (*Am J Public Health.* 1994;84:859–860)

Introduction

During the recent epidemic of *Salmonella enteritidis,* New York State has been particularly affected, having been the site of some of the earliest, largest, and greatest number of outbreaks.[1–3] Between 1985 and 1993, New York documented 253 salmonella outbreaks, of which 89 (35%) were due to *S enteritidis* (Table 1). Of the latter, 47 (53%) were egg associated, and these outbreaks resulted in at least 2279 cases and 10 deaths. Concurrently, the number of total salmonella cases reported through the state's notifiable disease surveillance system for upstate New York (excluding New York City) increased from 1513 (14.5 per 100 000) in 1984 to more than 2366 (21.9 per 100 000) in 1993 (provisional data).

Surveillance Efforts

The large number of *S enteritidis* outbreaks made it likely that a number of sporadic cases were also associated with eggs.[1,2,4] Two reports in the literature from the United Kingdom supported that conclusion.[5,6] More recently, Hedberg et al.[7] showed that persons with sporadic cases of *S enteritidis* (odds ratio [OR] = 5.2) and *Salmonella typhimurium* (OR = 2.4) were more likely than controls to have consumed undercooked eggs or egg-containing foods. We conducted a limited pilot study of risk factors for sporadic *S enteritidis* cases identified in three New York counties (Erie, Onondaga, and Monroe). During an 8-month period in 1989, additional surveillance information was collected on all adults with reported non–outbreak-associated, culture-confirmed salmonella and campylobacter cases who could be interviewed within 14 days of illness onset. Data were obtained for 45 cases of *S enteritidis* and compared with those for 316 cases of campylobacteriosis and 156 cases of non–*S enteritidis* salmonellosis (55 *S typhimurium,* 21 *Salmonella heidelberg,* and 80 other salmonella species). Results showed an association between illness and egg consumption within 3 days of illness for cases of *S enteritidis* relative to cases of campylobacteriosis and a borderline association relative to non–*S enteritidis* salmonella cases (see Table 2).

Conclusions

These studies' findings indicate that the number of illnesses related to consumption of raw or undercooked eggs is far greater than recognized by outbreak surveillance alone. This helps explain the failure of current US Department of Agriculture control practices, which are based on testing of laying flocks only if they are conclusively linked to an outbreak and identified through a traceback or are found on a voluntary testing basis.[8,9]

In response to the increase in *S enteritidis,* the New York State Department of Health has adopted regulations requiring the refrigeration of eggs in food service establishments and the cooking of all eggs and foods that contain eggs to 145°F (unless prepared differently at the request of a consumer). It has also recommended that recipes containing raw eggs (e.g., mayonnaise, Caesar salad dressing, and eggnog) be considered high risk and that pasteurized eggs be used for these preparations as well as in institutions serving high-risk populations (e.g., hospitals and nursing homes). Eggs are a highly nutritious food but must be cooked thoroughly; consumption of raw eggs, like raw milk, raw shellfish, and raw meat, can be a health hazard. In the absence of an effective control program, the Depart-

The authors are with the Bureaus of Communicable Disease Control and Community Sanitation and Food Protection, New York State Department of Health, Albany, and the Department of Epidemiology, School of Public Health, State University of New York at Albany.

Requests for reprints should be sent to Dale L. Morse, MD, MS, Bureau of Communicable Disease Control, New York State Department of Health, Tower Bldg, Room 651, Albany, NY 12237.

This paper was accepted August 23, 1993.

TABLE 1—Outbreaks of *Salmonella enteritidis* Associated with Eggs: New York, 1985 through 1993

Year	Total No. of Salmonella Outbreaks	No. of Outbreaks Due to *S enteritidis*[a]	No. of *S enteritidis* Outbreaks Associated with Eggs	No. of *S enteritidis* Egg-Associated Cases	No. of *S enteritidis* Egg-Associated Deaths
1985	16	5	1	20	0
1986	24	5	2	36	0
1987	29	7	6	621	9
1988	28	10	7	90	0
1989	37	17	8	119	0
1990	26	11	2	29	0
1991	27	15	9	790	0
1992	34	12	8	557	0
1993[b]	30	7	4	17	1
Total	253	89[a] (35%)	47 (53%)	2279	10

[a]This does not include an additional 18 outbreaks (374 cases) of Group D salmonella, most of which were probably *S enteritidis*.
[b]Preliminary data.

TABLE 2—Egg Consumption by Case Subjects with *Salmonella enteritidis* vs Campylobacter and Other Salmonella Case Subjects: Sporadic Cases, Three New York Counties, May through December 1989

Comparison	*S enteritidis* Case Subjects Who Ate Eggs		Other Salmonella Case Subjects Who Ate Eggs		Odds Ratio (95% Confidence Interval)	*P*
	No.	%	No.	%		
S enteritidis vs campylobacter	22	49	99	31	2.1 (1.06, 4.19)	.02
S enteritidis vs other salmonellas	22	49	55	35	1.8 (0.85, 3.62)	.10

ment of Agriculture should reevaluate its policies and issue additional public warnings about raw egg consumption. □

References

1. St Louis ME, Morse DL, Potter ME, et al. The emergence of grade A eggs as a major source of *Salmonella enteritidis* infections: new implications for the control of salmonellosis. *JAMA.* 1988;259:2103–2107.
2. Centers for Disease Control. Increasing role of *Salmonella enteritidis* infections in the Northeast. *MMWR.* 1987;36:10–11.
3. Telzak EE, Budnick LD, Greenberg MS, et al. A nosocomial outbreak of *Salmonella enteritidis* infection due to the consumption of raw eggs. *N Engl J Med.* 1990;323:394–397.
4. Anonymous. *Salmonella enteritidis* phage type 4: chicken and egg. *Lancet.* 1988;2:720–722.
5. Coyle EF, Palmer SR, Ribeiro CD, et al. *Salmonella enteritidis* phage type 4 infection; association with hen's eggs. *Lancet.* 1988;2:1295–1296.
6. Cowden JM, Lynch D, Joseph CA, et al. Case-control study of infections with *Salmonella enteritidis* phage type 4 in England. *Br Med J.* 1989;299:771–773.
7. Hedberg CW, David MJ, White KE, MacDonald KL, Osterholm MT. Role of egg consumption in sporadic *Salmonella enteritidis* and *Salmonella typhimurium* infections in Minnesota. *J Infect Dis.* 1993;167:107–111.
8. Centers for Disease Control. Update: *Salmonella enteritidis* infections and shell eggs—United States, 1990. *MMWR.* 1990;39:909–912.
9. Centers for Disease Control. Outbreak of *Salmonella enteritidis* infection associated with consumption of raw shell eggs, 1991. *MMWR.* 1992;41:369–372.

The Reproducibility of Data from a Food Frequency Questionnaire among Low-Income Latina Mothers and Their Children

Charles E. Basch, PhD, Steven Shea, MD, and Patricia Zybert, PhD

ABSTRACT

We examined the 3-month and 1-year reproducibility of mothers' and children's dietary intakes of calories and 11 nutrients derived from oral administrations of a food frequency questionnaire to 166 Latina women. One-year correlations generally ranged from 0.40 to 0.55 for both mothers and children. For most nutrients, roughly half the children in the highest (lowest) quartile of intake at baseline were also in the highest (lowest) quartile 1 year later. The food frequency questionnaire has great potential for measuring typical nutrient intakes in these populations. (*Am J Public Health.* 1994;84: 861–864)

Introduction

Reliable, valid, and practical measures of typical diet are needed for public health research.[1] Dietary recalls may accurately estimate nutrient intakes,[2,3] but large day-to-day variations in diet[4] necessitate multiple labor-intensive administrations to measure individuals' typical intakes.[4–6] Single administrations of a food frequency questionnaire have yielded acceptably reliable and inexpensive estimates of typical intake for a range of nutrients in various contexts.[5] Thus, despite limitations,[5–8] food frequency questionnaires have become the primary tool for dietary measurement in epidemiological studies.[5] We report 3-month and 1-year reproducibility of mothers' and children's dietary intakes of calories and 11 nutrients derived from oral administrations of a food frequency questionnaire to 166 Latina women. Reproducibility of nutrient intake estimates derived from food frequency data has not been reported previously for these groups.

Methods

Subjects and Setting

Subjects were 166 Latina mothers participating in a longitudinal study of generally healthy, low-income families living in New York City (Table 1). Subjects reported on both their own and their children's dietary intake on at least three occasions, using a modified version of the Willett food frequency questionnaire. Details of recruitment and eligibility have been published previously.[4,9,10] Mean ages of the mothers and their children at baseline were 33.3 years (SD = 5.1) and 5.4 years (SD = 0.4), respectively. Half the children were female.

Measurements

Modified Willett food frequency questionnaires[5,11] were administered to mothers three times for their own intakes and three times for their children's intakes over a 1-year period. Bilingual interviewers read a list of foods to mothers and, in a closed-response format, recorded the frequency with which each food was consumed.[5,11] The list included ethnic foods not on the original questionnaire and omitted foods originally listed but never eaten by this population. Children's portions were reduced to be consistent with portions reported in 24-hour dietary recalls from the same population.[2,4] The questionnaire's original 1-year time referent was reduced to 6 months. Nutrient intakes were computed using a database and program provided to us by Willett (W. Willett, MD, DrPH, personal communication, 1986, updated 1989), which we altered to match our food lists.

Statistical Analysis

Test-retest Pearson correlations were computed, and values for mothers versus children and for 3 months versus 1 year were compared using the *z* test applied to Fisher's *z* transformed correlations.[12] Since caloric adjustment of nutrient intakes[5,13,14] did not systematically affect reproducibility for the 11 nutrients, only results for unadjusted nutrient intakes are presented (Table 2). Regression equations and correlations were calculated after logarithmic transformation of intakes. The proportions of mothers and children consistently classified into the lowest and highest quartiles of caloric and nutrient intake over 1 year were also determined.

Results

Table 2 presents untransformed mean intakes of calories and 11 nutrients for

Charles E. Basch and Patricia Zybert are with the Center for Health Promotion, Teachers College, Columbia University, New York, NY. Steven Shea is with the College of Physicians and Surgeons and the School of Public Health, Columbia University.

Requests for reprints should be sent to Charles E. Basch, PhD, Center for Health Promotion, Box 114, Teachers College, Columbia University, New York, NY 10027.

This paper was accepted September 20, 1993.

TABLE 1—Selected Characteristics of 166 Latina Mothers, New York City, 1986 through 1988

	No.	%
Was born outside the United States	154	92.8
Emigrated from the Dominican Republic	127	76.5
Is unemployed	162	97.6
Receives Medicaid	144	86.7
Has less than 12 years of education	141	84.9
Uses Spanish as preferred language	158	95.0
Is a single head of household	102	61.4

TABLE 2—Mean Daily Caloric and Nutrient Intakes for 166 Latino Mothers and Children, Derived from a Single, Orally Administered Food Frequency Questionnaire, New York City, 1986 through 1988

	Mothers		Children	
	Mean	Standard Deviation	Mean	Standard Deviation
Calories	2 043.2	774.6	2 565.7	759.2
Protein, g	83.7	27.1	86.2	21.1
Total fat, g	60.7	27.1	93.1	33.2
Total fat, % of calories	26.5	6.8	32.5	5.4
Saturated fat, g	22.4	10.4	36.7	12.1
Saturated fat, % of calories	9.7	2.4	12.9	2.3
Polyunsaturated fat, g	10.8	4.9	14.9	6.4
Cholesterol, mg	348.4	170.8	387.1	126.3
Cholesterol/1000 kcal	174.6	66.1	154.0	41.6
Carbohydrates, g	304.6	123.8	370.8	123.7
Sucrose, g	56.7	26.3	89.8	43.8
Crude fiber, g	6.7	2.7	6.0	2.2
Vitamin A, IU	19 183	11 993	12 446	6 859
Vitamin B6, mg	2.65	1.18	3.08	1.47
Vitamin C, mg	241.1	171.8	277.1	199.6

TABLE 3—1-Year and 3-Month Test–Retest Correlations for Caloric and Nutrient Intakes for 166 Latino Mothers and Children

	Mothers				Children			
	3-Month Interval[a]		1-Year Interval[b]		3-Month Interval[c]		1-Year Interval[d]	
	r	(95% CI)	*r*	(95% CI)	*r*	(95% CI)	*r*	(95% CI)
Calories	.59	(.48, .68)	.51	(.39, .61)	.53	(.42, .63)	.46	(.33, .57)
Protein, g	.49	(.36, .60)	.42	(.29, .54)	.49	(.36, .60)	.40	(.27, .52)
Total fat, g	.60	(.50, .69)	.47	(.35, .58)	.56	(.45, .66)	.47	(.34, .58)
Total fat, % of calories	.69	(.50, .76)	.56	(.44, .65)	.56	(.45, .66)	.41	(.27, .53)
Saturated fat, g	.56	(.44, .65)	.46	(.33, .57)	.46	(.33, .57)	.38	(.24, .50)
Saturated fat, % of calories	.54	(.42, .64)	.38	(.25, .51)	.36	(.22, .49)	.21	(.06, .35)
Polyunsaturated fat, g	.64	(.54, .72)*	.37	(.23, .49)	.53	(.41, .63)	.57	(.46, .67)
Cholesterol, mg	.51	(.39, .62)	.36	(.22, .49)	.53	(.41, .63)	.38	(.24, .50)
Cholesterol/1000 kcal	.31	(.17, .44)	.29	(.15, .43)	.46	(.33, .57)	.36	(.22, .48)
Carbohydrates, g	.59	(.48, .68)	.54	(.42, .64)	.53	(.42, .63)	.42	(.29, .54)
Sucrose, g	.46	(.33, .58)	.48	(.36, .59)	−.06	(.00, .21)	.06	(.00, .21)**
Crude fiber, g	.61	(.51, .70)	.56	(.44, .65)	.61	(.51, .70)	.53	(.41, .63)
Vitamin A, IU	.58	(.47, .67)	.46	(.33, .57)	.53	(.41, .63)	.55	(.44, .65)
Vitamin B6, mg	.45	(.31, .56)	.35	(.21, .47)	.38	(.25, .51)	.35	(.20, .47)
Vitamin C, mg	.53	(.41, .63)	.43	(.30, .55)	.31	(.16, .44)	.23	(.08, .37)

Note. CI = confidence interval.
[a]Mean interval = 84 days (SD = 26 days), 25th percentile = 69 days, 75th percentile = 95 days.
[b]Mean interval = 364 days (SD = 73 days), 25th percentile = 294 days, 75th percentile = 428 days.
[c]Mean interval = 86 days (SD = 24 days), 25th percentile = 72 days, 75th percentile = 98 days.
[d]Mean interval = 374 days (SD = 54 days), 25th percentile = 338 days, 75th percentile = 413 days.
*$z = 3.4$, $P < .01$, for mothers' 1-year vs mothers' 3-month interval.
**$z = 4.3$, $P < .01$, for children's 1-year vs mothers' 1-year interval.

mothers and children at baseline. Nutrient intakes derived from multiple 24-hour recalls on the same children[10] suggest substantial inflation in the food frequency–derived estimates.[15,16] Table 3 presents 3-month and 1-year test-retest correlations. Disregarding densities, 3-month correlations among mothers ranged from .45 for vitamin B6 to about .60 for polyunsaturated fat, crude fiber, total fat, calories, carbohydrates, and vitamin A. Among the children, the 3-month correlations ranged from −.06 for sucrose to more than .50 for crude fiber, total fat, carbohydrates, calories, polyunsaturated fat, cholesterol, and vitamin A. One-year correlations among mothers ranged from .35 for vitamin B6 to more than .5 for crude fiber, carbohydrates, and calories. The 1-year correlations for children were similar, except that children's sucrose intake was less stable at 1 year than their mothers' ($z = 4.3, P < .01$). Reproducibility over the shorter interval tended to be higher for both mothers and children. For most nutrients, approximately half the mothers and children in the highest (lowest) quartile of intake at baseline were also in the highest (lowest) quartile 1 year later (Table 4).

Discussion

Latino people and groups with low income have been designated as special target populations in many diet-related health objectives for the year 2000.[1] Compared with higher income non-Hispanic Whites, low-income Hispanics have higher triglycercide levels,[17,18] lower high-density lipoprotein cholesterol levels among women,[17] increased prevalence of diabetes[19–23] and obesity,[17,23,24] and higher Framingham cardiovascular risk scores.[24] Reliable, valid, and practical dietary measures for use in these populations are needed to monitor diet-related risk reduction goals.[1,25–31]

Our 1-year test–retest correlations for mothers' intakes based on a single, orally administered food frequency questionnaire generally ranged between .40 and .55. In Willett's study of 173 women in the Nurses Health Study, 1-year

Pearson correlations ranged from .50 to .65 for most unadjusted nutrients.[11] However, ethnic, social, and behavioral differences should be considered when comparing these studies. The acculturation process may produce greater instability of dietary intake among our immigrant population. Low-income non–English-speaking women may have different body weight norms[32] and correspondingly different levels of attentiveness to the details of their diet than the health professionals studied by Willett et al.[11] Also, the women in our study may have been less prepared to respond to the structured questions typical of standardized data collection protocols. Finally, our changing of the time referent from 1 year to 6 months may have reduced reproducibility by introducing seasonal variation.

Some studies support the value of surrogate reports for measuring diet[2,3] while others document shortcomings.[33,34] We found that a single food frequency questionnaire administered orally to mothers provided 1-year and 3-month test-retest correlations for children's intakes generally ranging between .40 and .55 and between .45 and .55, respectively. Treiber et al. observed 1-week test-retest correlations for 55 children from middle socioeconomic class families that were generally between .50 and .70[35]; their higher correlations are compatible with the shorter time interval. Eck et al. report average 4- and 8-month test-retest correlations for 108 young children from intact families generally in the range of .60 to .65.[36] Our data[10] as well as those of Treiber et al.[35] indicate that food frequency questionnaires yield inflated estimates of children's absolute intakes.

While a single 24-hour recall may provide valid intake estimates for a single day, these estimates are not reproducible because of day-to-day variations in an individual's diet.[4] Thus, a single recall does not measure typical diet. In contrast, food frequency questionnaires may yield inflated estimates of absolute intakes, but they are presumably reproducible because they reflect typical diet. This paper does not address the validity of food frequency–derived estimates. Further research is needed to calibrate the food frequency to yield more accurate data.

Several limitations of this study restrict the scope of conclusions that can be drawn. First, we cannot generalize to nutrients we did not examine. Second, we did not identify or control variations in food availability or accessibility, either of which might have confounded our results.

TABLE 4—Proportions of 166 Mothers and Children Classified into Lowest (Highest) Quartile of Caloric and Nutrient Intakes at Baseline Who Remained in Lowest (Highest) Quartile or Half 1 Year Later

	Mothers				Children			
	Lowest Baseline		Highest Baseline		Lowest Baseline		Highest Baseline	
	Lowest Quartile	Lower Half	Highest Quartile	Higher Half	Lowest Quartile	Lower Half	Highest Quartile	Higher Half
Calories	.51	.80	.51	.76	.46	.80	.44	.73
Protein, g	.39	.76	.49	.76	.54	.80	.46	.66
Total fat, g	.56	.78	.54	.78	.41	.80	.54	.80
Total fat, % of calories	.54	.78	.46	.61	.32	.71	.59	.83
Saturated fat, g	.54	.83	.46	.78	.49	.80	.41	.71
Saturated fat, % of calories	.49	.73	.41	.68	.49	.66	.29	.68
Polyunsaturated fat, g	.41	.78	.49	.80	.51	.80	.51	.80
Cholesterol, mg	.34	.63	.49	.68	.41	.76	.46	.71
Cholesterol/1000 kcal	.51	.68	.37	.63	.46	.73	.41	.59
Carbohydrates, g	.63	.83	.43	.79	.49	.85	.39	.68
Sucrose, g	.54	.83	.27	.63	.46	.78	.44	.68
Crude fiber, g	.54	.76	.51	.85	.59	.88	.56	.83
Vitamin A, IU	.44	.71	.46	.61	.49	.83	.51	.80
Vitamin B6, mg	.36	.67	.49	.73	.40	.67	.39	.76
Vitamin C, mg	.51	.68	.41	.63	.41	.66	.32	.59

Third, we did not have a gold standard and could not distinguish true change in diet from measurement error. Fourth, we could not determine the specific sources of measurement errors.

We view our results as encouraging. Our 1-year reproducibility of food frequency–derived nutrient intakes for both mothers and children was only slightly lower than the published reproducibility[11] at which the food frequency questionnaire has been found to be a useful tool for epidemiological research.[5,11] When using the current version of the Willett food frequency questionnaire among low-income non–English-speaking adults, we recommend oral administration coupled with enhancement of stability through averaging of the results from multiple administrations. Analyses involving children's absolute intakes should be avoided. Following these guidelines, we have found relationships between ordinal classification by children's saturated fat intake and serum lipids[9] and between mothers' food choice beliefs and children's nutrient intakes.[37] Both of these results were confirmed using nutrient intakes derived from multiple 24-hour recalls.[9,37] Research is needed to identify the sources of instability in food frequency measurements and, for children, to determine the factors that result in inflated intake estimates. □

Acknowledgments

This research was supported in part by grants from the National Heart, Lung, and Blood Institute (HL-35189 and HL-49508) and the Cancer Research Foundation of America. At the time of the study, Steven Shea was a Florence Irving Assistant Professor.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50212.
2. Basch CE, Shea S, Arliss A, et al. Validation of mothers' reports of dietary intake by four to seven year-old children. *Am J Public Health.* 1990;80:1314–1317.
3. Baranowski T, Sprague D, Baranowski JH, Harrison JA. Accuracy of maternal dietary recall for preschool children. *J Am Diet Assoc.* 1991;91:669–674.
4. Stein AD, Shea S, Basch CE, Contento IR, Zybert P. Variability and tracking of nutrient intakes of preschool children based on multiple administrations of the 24-hour dietary recall. *Am J Epidemiol.* 1991;134:1427–1437.
5. Willett WC. *Nutritional Epidemiology.* New York, NY: Oxford University Press; 1990.
6. Rimm EB, Giovannucci EL, Stampfer MJ, et al. Authors' response to "invited commentary: some limitations of semiquantita-

tive food frequency questionnaires." *Am J Epidemiol.* 1992;135:1133–1136.
7. Sempos CT. Invited commentary: some limitations of semiquantitative food frequency questionnaires. *Am J Epidemiol.* 1992;135:1127–1132.
8. Briefel RR, Flegel KM, Winn DM, et al. Assessing the nation's diet: limitations of the food frequency questionnaire. *J Am Diet Assoc.* 1992;92:959–962.
9. Shea S, Basch CE, Irigoyen M, et al. Relationships of dietary fat consumption to serum total and low-density lipoprotein cholesterol in Hispanic preschool children. *Prev Med.* 1991;20:237–249.
10. Stein AD, Shea S, Basch CE, et al. Consistency of the Willett semiquantitative food frequency questionnaire and 24-hour dietary recalls in estimating nutrient intakes of preschool children. *Am J Epidemiol.* 1992;135:667–677.
11. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol.* 1985;122:51–65.
12. Snedecor GW, Cochran WG. *Statistical Methods.* 6th ed. Ames, Iowa: Iowa State University Press; 1967.
13. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analysis. *Am J Epidemiol.* 1986;124:17–27.
14. Willett WC. Total energy intake and nutrient composition: dietary recommendations for epidemiologists. *Int J Cancer.* 1990;46:770–771.
15. Nationwide food consumption survey, continuing survey of food intakes by individuals: women 19–50 years and their children 1–5 years, 4 days. Hyattsville, Md: US Dept of Agriculture; 1987. Human Nutrition Information Service NFCS, CSFII, Report 85-4.
16. Kimm SYS, Gergen PJ, Malloy M, Dresser C, Carroll M. Dietary patterns of US children: implications for disease prevention. *Prev Med.* 1990;19:432–434.
17. Stern MP, Rosenthal M, Haffner SM, Hazuda HP, Franco LJ. Sex differences in the effects of sociocultural status on diabetes and cardiovascular risk factors in Mexican Americans: the San Antonio Heart Study. *Am J Epidemiol.* 1984;120:834–851.
18. Stern MP, Gaskill SP, Allen CR, et al. Cardiovascular risk factors in Mexican Americans in Laredo, Texas, I: prevalence of overweight and diabetes and distributions of serum lipids. *Am J Epidemiol.* 1981;113:546–555.
19. Stern MP, Haffner SM. Type II diabetes and its complications in Mexican Americans. *Diabetes Metab Rev.* 1990;6:29–45.
20. Samet JM, Coultas DB, Howard CA, Skipper BJ, Hanis CL. Diabetes, gall bladder disease, obesity and hypertension among Hispanics in New Mexico. *Am J Epidemiol.* 1988;128:1302–1311.
21. Hamman RF, Marshall JA, Baxter J, et al. Methods and prevalence of non–insulin-dependent diabetes mellitus in a biethnic population: the San Luis Valley Diabetes Study. *Am J Epidemiol.* 1989;129:295–311.
22. Harris MI. Epidemiologic correlates of NIDDM in Hispanics, Whites and Blacks in the US population. *Diabetes Care.* 1991; 14(suppl 3):639–648.
23. Diehl AK, Stern MP. Special health problems of Mexican Americans: obesity, gall bladder disease, diabetes mellitus, and cardiovascular disease. *Adv Intern Med.* 1989;34:79–95.
24. Mitchell BD, Stern MP, Haffner SM, Hazuda HP, Patterson JK. Risk factors for cardiovascular mortality in Mexican Americans and non-Hispanic Whites: San Antonio Heart Study. *Am J Epidemiol.* 1990;131: 423–433.
25. Committee on Nutrition, American Academy of Pediatrics. Statement on cholesterol. *Pediatrics.* 1992;90:469–473.
26. *National Cholesterol Education Program: Report of the Expert Panel on Blood Cholesterol Levels in Children and Adolescents.* Washington, DC: US Public Health Service; 1991. National Institutes of Health publication 91-2732.
27. National Cholesterol Education Program. Report of the Expert Panel on Population Strategies for Blood Cholesterol Reduction: Executive Summary. *Arch Intern Med.* 1991;151:1071–1084.
28. National Institutes of Health Consensus Development Conference statement. Lowering blood cholesterol to prevent heart disease. *JAMA.* 1985;253:2080–2086.
29. National Research Council, Committee on Diet and Health. *Diet and Health: Implications for Reducing Chronic Disease Risk.* Washington, DC: National Academy Press; 1989.
30. *Nutrition and Your Health: Dietary Guidelines for Americans.* 3rd ed. Washington, DC: US Dept of Agriculture and US Dept of Health and Human Services; 1990.
31. *The Surgeon General's Report on Nutrition and Health.* Washington, DC: US Dept of Health and Human Services; 1988.
32. Dawson DA. Ethnic differences in female overweight: data from the National Health Interview Survey. *Am J Public Health.* 1988;78:1326–1329.
33. Humble CG, Samet JM, Skipper BE. Comparison of self- and surrogate-reported dietary information. *Am J Epidemiol.* 1984;119:86–98.
34. Marshall J, Priore R, Haughey B, et al. Spouse-subjects interviews and the reliability of diet studies. *Am J Epidemiol.* 1980;112: 675–683.
35. Treiber FA, Leonard SB, Frank GC, et al. Dietary assessment instruments for preschool children: reliability of parental surrogate responses to the 24-hour recall and a food frequency questionnaire. *J Am Diet Assoc.* 1990;90:814–820.
36. Eck LH, Klesges RC, Hanson CL. Reporting retrospective dietary intake by food frequency questionnaire in a pediatric population. *J Am Diet Assoc.* 1991;91:606–608.
37. Contento IR, Basch CE, Shea S, et al. Relationship of mothers' food choice criteria to food intake of preschool children: identification of family subgroups. *Health Educ Q.* 1993;20:243–259.

A2623-1

American Journal of Public Health

June 1994, Vol. 84, No. 6 Established 1911

RTER
IONAL INFORMATION SVC
AST-WEST HIGHWAY
ESDA MD 20814
9375

APHA
American Public Health Association

Featuring Disadvantage in Health

Editorials

888 Formulating Public Health Policy: The Case of AIDS *K. M. Gebbie*
890 The Urgent Need for a Vaginal Microbicide in the Prevention of HIV Transmission *M. Potts*
892 Elucidating the Relationships between Race, Socioeconomic Status, and Health *G. Pappas*
893 Why a Rising Tide Doesn't Lift All the Boats: Medicaid and Medical Care for Children *G. A. Silver*

Commentary

895 AIDS Prevention and Cultural Sensitivity: Are They Compatible? *R. Bayer*

Public Health Policy Forum

899 Rationing or Rationalizing Children's Medical Care: Comparison of a Medicaid HMO with Fee-for-Service Care *J. Mauldon, A. Leibowitz, J. L. Buchanan, C. Damberg, and K. A. McGuigan*
905 The Role of State Health Departments in Formulating Policy: A Survey on the Case of AIDS *L. Robins and C. Backstrom*

Articles

VARIOUS TOPICS

910 The Use of Nonoxynol-9 for Protection against Cervical Gonorrhea *S. S. Weir, P. J. Feldblum, L. Zekeng, and R. E. Roddy*
915 Are Bisexually Identified Men in San Francisco a Common Vector for Spreading HIV Infection to Women? *M. L. Ekstrand, T. J. Coates, J. R. Guydish, W. W. Hauck, L. Collette, and S. B. Hulley*
920 Correlates of Needle Sharing among Injection Drug Users *W. Mandell, D. Vlahov, C. Latkin, M. Oziemkowska, and S. Cohn*
924 The Effects of Early Education Intervention on Maternal Employment, Public Assistance, and Health Insurance: The Infant Health and Development Program *J. Brooks-Gunn, M. C. McCormick, S. Shapiro, A. A. Benasich, and G. W. Black*

DISADVANTAGE IN HEALTH

932 The Size of Mortality Differences Associated with Educational Level in Nine Industrialized Countries *A. E. Kunst and J. P. Mackenbach*
938 Why Did Black Life Expectancy Decline from 1984 through 1989 in the United States? *K. D. Kochanek, J. D. Maurer, and H. M. Rosenberg*
945 The Effects of Occupational Class Transitions on Hypertension: Racial Disparities among Working-Age Men *N. J. Waitzman and K. R. Smith*
951 Diagnoses, Symptoms, and Attribution of Symptoms among Black and White Inpatients Admitted for Coronary Heart Disease *J. M. Raczynski, H. Taylor, G. Cutter, M. Hardin, N. Rappaport, and A. Oberman*
957 Do Blacks and Whites Differ in Their Use of Health Care for Symptoms of Coronary Heart Disease? *S. L. Crawford, S. A. McGraw, K. W. Smith, J. B. McKinlay, and J. E. Pierson*
965 Acute Chest Pain in African Americans: Factors in the Delay in Seeking Emergency Care *K. Ell, L. J. Haywood, E. Sobel, M. deGuzman, D. Blumfield, and J.-P. Xiang*
971 Behavioral Risk Factors: A Comparison of Latinos and Non-Latino Whites in San Francisco *E. J. Pérez-Stable, G. Marín, and B. VanOss Marín*

(continued on page 881)

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

June 1994, Vol. 84, No. 6 ISSN: 0090-0036 Established 1911

977 Social Influences on the Sexual Behavior of Youth at Risk for HIV Exposure *D. Romer, M. Black, I. Ricardo, S. Feigelman, L. Kaljee, J. Galbraith, R. Nesbit, R. C. Hornik, and B. Stanton*

986 The Adequacy of Prenatal Care and Incidence of Low Birthweight among the Poor in Washington State and British Columbia *S. J. Katz, R. W. Armstrong, and J. P. LoGerfo*

Public Health Briefs

VARIOUS TOPICS

992 Smoking and Mortality among Residents of a California Retirement Community *A. Paganini-Hill and G. Hsu*

995 The Utility of HMO Data for the Surveillance of Chronic Diseases *D. L. Nordstrom, P. L. Remington, and P. M. Layde*

998 Popular Films Do Not Reflect Current Tobacco Use *A. R. Hazan, H. L. Lipton, and S. A. Glantz*

DISADVANTAGE IN HEALTH

1000 The Resurgence of Tuberculosis in New York City: A Mixed Hierarchically and Spatially Diffused Epidemic *D. Wallace*

1003 Temporal Trends in the Socioeconomic Gradient for Breast Cancer Mortality among US Women *D. K. Wagener and A. Schatzkin*

1007 Racial Differences in Cervical Cancer Mortality in Chicago *E. J. Samelson, M. A. Speers, R. Ferguson, and C. Bennett*

1010 Access to Family Planning Services and Health Insurance among Low-Income Women in Arizona *B. Kirkman-Liff and J. J. Kronenfeld*

1013 Sex Bias in the Management of Coronary Artery Disease in Quebec *W. D'Hoore, C. Sicotte, and C. Tilquin*

1015 Health Insurance Coverage among Persons with AIDS: Results from a Multistate Surveillance Project *T. Diaz, S. Y. Chu, L. Conti, B. L. Nahlen, B. Whyte, E. Mokotoff, A. Shields, P. J. Checko, M. Herr, Q. Mukhtar, C. A. Rietmeijer, A. Levy, P. Hermann, and J. W. Buehler*

1018 The Reliability of Racial Classifications in Hospital Discharge Abstract Data *J. Blustein*

Public Health Then and Now

1022 The Trials of Hanna Porn: The Campaign to Abolish Midwifery in Massachusetts *E. R. Declercq*

Notes from the Field

1029 A Homeless Teen Pregnancy Project: An Intensive Team Case Management Model *D. Borgford-Parnell, K. R. Hope, and R. W. Deisher*

1030 A Population-Based Screening Mammography Program in Saskatchewan *L. Edouard, L. Eberts, and J. Baldwin*

Other Journal Departments

882 APHA Masthead

884 What *AJPH* Authors Should Know

886 Advertisers' Index

888 Journal Masthead

1032 Letters to the Editor

1039 Book Corner

1042 Annual Meeting Registration Form and Housing Information

1048 Job Opportunities

American Public Health Association

American Journal of Public Health

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 15th St, NW, Washington, DC 20005; (202) 789-5600; TDD (202) 789-5673. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*

Helen Rodriguez-Trias, MD, *Immediate Past President*

Caswell A. Evans, Jr, DDS, MPH, *President-Elect*

Leonard M. Schuman, MD, *Vice-President (USA)*

John E. F. Hastings, MD, MPH, *Vice-President (Canada)*

Oscar Feo, MD, MSc, *Vice-President (Latin America)*

Jay H. Glasser, PhD, *Treasurer*

Allan G. Rosenfield, MD, *Chair, Executive Board*

Shirley F. Randolph, MSPH, *Speaker, Governing Council*

Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*

Carol Easley Allan, PhD, RN (1997)

Caswell A. Evans, Jr, DDS, MPH, *President-Elect*

Patricia P. Evans, MPH (1994)

Eugene Feingold, PhD, JD, *President*

Jay H. Glasser, PhD, *Treasurer*

Alan R. Hinman, MD, MPH (1995)

Margo D. Kerrigan, MPH (1995)

Joanne Lukomnik, MD (1996), *Vice-Chair*

Steven Uranga McKane, DMD, MPH (1995)

Carlos W. Molina, EdD (1997)

Linda Rae Murray, MD, MPH (1996)

Hila Richardson, DrPH, MPH (1996)

Helen Rodriguez-Trias, MD, *Immediate Past President*

Reed V. Tuckson, MD (1994)

Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*

Frances Kissling, *Chair, Action Board*

George E. Pickett, MD, MPH, *Speaker, Program Development Board*

Shirley F. Randolph, MSPH, *Speaker, Governing Council*

The *American Journal of Public Health* (*AJPH*), published monthly, is the official Journal of the American Public Health Association, Inc, 1015 15th St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005. *AJPH* allotted $40; *The Nation's Health* allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters. Printed in the United States.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on the first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, *AJPH* Reprints, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the Journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The *American Journal of Public Health* is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

Public Health Policy Forum

Rationing or Rationalizing Children's Medical Care: Comparison of a Medicaid HMO with Fee-for-Service Care

Jane Mauldon, PhD, Arleen Leibowitz, PhD, Joan L. Buchanan, PhD, Cheryl Damberg, MPH, and Kimberly A. McGuigan, MS

ABSTRACT

Objectives. This paper examines how medical care obtained by children enrolled in a Medicaid health maintenance organization (HMO) differs from that obtained by similar children who receive care from fee-for-service Medicaid providers.

Methods. In a randomized trial, some Medicaid households were assigned to remain in a traditional fee-for-service arrangement and others were randomly selected to join a Medicaid prepaid plan (an HMO). Participating households recorded data on children's health status and use of medical care.

Results. The prepaid plan members and the fee-for-service recipients received equivalent numbers of checkup visits, but the children in the prepaid plan made significantly fewer acute care visits. This plan appeared to target its services to children with the greatest health care needs. However, the content of health visits in the two systems did not differ, nor did prepaid and fee-for-service enrollees use the emergency room differently.

Conclusions. It is possible to design a Medicaid HMO that achieves financial savings without reducing services to the most vulnerable patients. However, these findings alone do not provide a basis for widespread policy change in the direction of Medicaid HMOs. Further research is needed to establish whether the children treated in the HMO differed in health outcomes from those treated by fee-for-service care. (*Am J Public Health.* 1994;84: 899–904)

Introduction

In recent years, policymakers and the research community have been discussing structural reforms for the Medicaid program that would control costs without creating further access barriers to patients. Enrolling Medicaid beneficiaries in health maintenance organizations (HMOs) instead of paying for their care on a fee-for-service basis is one of the most common proposals. However, it is important to determine whether the HMO's rationing of care further reduces access to needed services, or whether HMOs reduce costs by "rationalizing" service delivery to provide care for those with greatest needs while eliminating unnecessary services.

Given the considerable policy interest in HMOs, the Robert Wood Johnson Foundation, under its Program for Prepaid Managed Health Care, funded 13 demonstration projects in 1984 to expand the use of case management and capitation within the Medicaid population in 10 states across the country. The purpose of these experiments was to study the consequences of substituting HMO-type plans for traditional fee-for-service arrangements in the Medicaid program. The Health Care Financing Administration and the National Governors' Association joined the Robert Wood Johnson Foundation in sponsoring the RAND Corporation to evaluate these programs. This paper, part of the evaluation, reports data on children's use of outpatient care from one of the experimental sites.

Fee-for-Service Medicaid vs HMOs

The constraints and incentives that operate in fee-for-service Medicaid programs can generate inappropriate patterns of care for enrollees. One problem is that low Medicaid reimbursement levels discourage doctors from treating Medicaid patients, so many enrollees have limited or no access to private physician care. Consequently, many Medicaid patients seek care in hospital emergency rooms.[1–4] However, hospital emergency rooms are inappropriate sources of routine care because they provide episodic, high-cost services. Children who use such facilities as a regular place of care may not receive routine checkups and preventive care. Further, they may not receive *timely* care for incipient problems that would be noted and possibly treated during a checkup by a personal physician. Work by Kasper confirms that young children get more health care when they have a regular provider and not just a regular place of care.[5] HMOs have the potential to move the routine care of Medicaid beneficiaries out of hospital emergency rooms and into private physicians' offices, health centers, or outpatient clinics.

A second problem is that fee-for-service systems may induce excessive visits to private physicians while still not guaranteeing that patients receive all the medical care they need. When physicians are paid only for actual patient visits (and not, e.g., for telephone consultations), they may try to achieve adequate total reimbursement by encouraging multiple short visits for

Jane Mauldon is with the Graduate School of Public Policy, University of California at Berkeley. The other authors are with the RAND Corporation, Santa Monica, Calif.

Requests for reprints should be sent to Arleen Leibowitz, PhD, RAND, 1700 Main St, Santa Monica, CA 90407-2138.

Note. The views expressed here are the authors' and should not be construed as reflecting those of the funding agencies.

This paper was accepted November 29, 1993.

simple problems. In contrast, HMOs and other providers under capitated payment systems seek to offer care in the most cost-effective ways, which may mean replacing patient visits with telephone consultations or using fewer visits to treat a problem. If this is in fact the outcome, using HMOs for Medicaid clients could improve care and reduce unnecessary use. However, prepaid plans present incentives to reduce use generally, which may discourage medically necessary patient visits and lead to underprovision of needed services.

Experimental Reforms in Medicaid

Policy interest in Medicaid HMOs has been translated into legislative action on several occasions. Section 2178 of the Omnibus Budget Reconciliation Act of 1981 broadened states' opportunities to contract with existing HMOs for Medicaid services, while section 1915 invited states to experiment with both new forms of prepaid care and new forms of case management for Medicaid enrollees (under waivers from the existing Medicaid legislation).

Programs established under section 1915 were of particular interest to the policy research community. Many of these programs were included in the Medicaid Competition Demonstration Project, a multisite research project to assess opportunities for Medicaid reform; findings from the project provide the backdrop for our findings reported below.

A quasi-experimental evaluation of programs in Monterey and Santa Barbara, Calif, and in Kansas City, Mo (which were all "mandatory"—i.e., beneficiaries did not have the option of remaining in the fee-for-service system), and of a voluntary program in New Jersey, found strong and pervasive effects of gatekeeping mechanisms on the likelihood of an emergency room visit. In all four programs, the gatekeeper (an individual physician or primary care organization) was required to provide all primary care. Prior authorization by the gatekeeper was required before enrollees could receive nonurgent emergency room care, inpatient care, or services from other physicians. The study compared use of the emergency room among stratified random samples of Medicaid beneficiaries drawn from the experimental counties and nearby comparison counties. Much smaller proportions of both children and adults used emergency room services at least once in the experimental sites. However, numbers of emergency room visits for persons with at least one such visit were less substantially affected by the plans.[6]

Bonham and Barber examined the effect of a mandatory HMO Medicaid program on usage through in-depth interviews with randomly chosen Medicaid recipients, both prior to the start of the program and 1 year into it.[7] They found a 40% reduction in (self-reported) use of hospital emergency rooms, apparently owing to less use of the facilities for nonurgent care. There was no change, however, in the rate of hospital inpatient stays, ambulatory care, or prescribed drugs.

Two of the nine programs for which there are published estimates of program effects failed to reduce emergency room use, in each case for quite specific and readily identifiable reasons. In one program, pediatric emergency room use had been relatively low at the outset (i.e., under the fee-for-service system), so there was not a lot of room for improvement. In addition, some of the emergency room use was unapproved by the case manager and consequently did not have to be paid for by Medicaid.[8] The other plan that failed to change patterns of emergency room use was hospital based, so the hospital may not have faced strong financial incentives to shift care from the emergency room to the HMO clinic because it ran both.[9]

Most evaluations of non-Medicaid HMOs have found that HMOs increase the number of outpatient visits relative to fee-for-service care.[10-12] This is because HMO patients face minimal or no cost sharing while fee-for-service patients must pay at least part of the cost for outpatient visits. For Medicaid patients, the situation differs. Neither the HMO nor fee-for-service Medicaid requires patient cost sharing. Thus, if Medicaid patients use more visits in HMOs, this would reflect improved physical access rather than greater affordability.

HMOs may not provide an ideal alternative for Medicaid patients since studies suggest that low-income patients may not fare as well in HMOs as middle-income enrollees. Ware et al. show that, in a randomized trial, low-income adults with preexisting health problems reported more serious health problems and more bed days per year when they were assigned to an HMO rather than to fee-for-service care.[13] Although a study of children in the same HMO assignment revealed no significant negative health effects for those of either low or middle income,[14] the Ware finding suggests the importance of examining this issue for Medicaid HMO enrollment.

The Children's Medicaid Program in New York compared use of services for children treated in one of three systems: (1) traditional fee-for-service Medicaid; (2) a fee-for-service Medicaid group with case management and augmented provider payments; and (3) a capitated, case-managed traditional HMO group with risk sharing for referred services. In their evaluation of this system, Hohlen et al. report that children in the capitated system received at least as much care as regular Medicaid children through the combination of physician office and outpatient clinic visits.[15]

While these studies suggest quite strongly that HMOs reduce emergency room use and may under some circumstances increase preventive use, they still leave many questions unanswered. In particular, the studies do not indicate which types of emergency room use are constrained by gatekeeping systems (although Bonham and Barber[7] provide some data on this) and which types of patients are most inclined to change their behavior when they enter an HMO. It does not appear that the studies controlled for such confounding factors as overall health status and existing health problems. Nor do these studies reveal whether or how children's patterns of care seeking change when they shift from fee-for-service to HMO systems. For example, is more change seen in routine visits or in acute visits? The study reported in this paper begins to fill some of these gaps in knowledge.

Our investigation examines three possible benefits of HMOs over fee-for-service Medicaid: (1) that HMO enrollees use emergency rooms less because the HMO makes primary care services more available; (2) that patients in HMOs have more checkup visits (partly because office-based care is available to them); and (3) that because preventive care is available, patients make fewer acute care visits. We test these hypotheses for pediatric services only.

Methods

Data

The data used here are from a randomized clinical trial in which some Medicaid households were assigned to remain in a traditional fee-for-service arrangement, others were randomly selected to join a Medicaid HMO, and still

others self-selected into either fee-for-service or HMO care.

The study HMO was situated in a hospital that was the primary Medicaid health provider for the area. Patients' access to services was better under the HMO than under fee-for-service Medicaid because patients had a case manager and could make appointments for specific times rather than being scheduled for 3-hour "block appointments."

Focusing exclusively on use of care by children, this investigation uses three types of data: characteristics of children and their families (obtained from baseline interviews), Medicaid eligibility and health plan enrollment data (obtained from administrative records), and use of care in a given 2-month period (recorded in diaries administered at baseline and 2 and 4 months thereafter). The use data come from mothers' diary reports of their children's health care use. The diaries are relied upon because, unlike the claims files, they distinguish well care visits from acute care visits. The 1867 children's diaries available for this analysis cover children aged 0 to 14.

TABLE 1—Outcome Measures, Demographic Characteristics, and Plan Status Variables

	Mean	SD
Dependent variables		
Percentage with checkup visit	24.7	0.43
Percentage with acute care visit	18.7	0.39
Percentage with emergency room visit, given any visit	12.7	0.33
Independent variables		
Child's age, y	6.2	3.74
Plan status, %		
Enrollees in HMO plan	51.0	0.50
Randomly assigned	53.7	0.50
In HMO plan and randomly assigned	21.1	0.41
Questionnaire timing, %		
Summer (April to September)	32.5	0.47
First wave	21.0	0.41
Second wave	66.5	0.47
Third wave	12.5	0.33
White (i.e., non-Hispanic, non-Black), %	13.0	0.34
Number of health problems reported per child at start of experiment		
Allergies	0.20	0.49
Serious health problems	0.01	0.09
Other health problems	0.30	0.64

Note. No. of diaries = 1867; no. of individual children = 1685.

The sampling frame was constructed from Medicaid eligibility files, including only those families eligible for Medicaid by virtue of eligibility for Aid to Families with Dependent Children. From the set of eligible families who currently received care in the fee-for-service Medicaid system, two subsamples were drawn. One was labeled the random assignment group and the other was the self-select group. Families in the first group were asked to agree to be randomly assigned to either receive their care in an HMO or continue to receive fee-for-service care. Thus, these families were comparable in that they all agreed, in principle, to enroll in an HMO if they were randomly assigned to do so. The actual plan assignment was not disclosed to random assignment families until after they had agreed to participate in the study. To determine whether families who agreed to randomization differed systematically from the average family enrolled in Medicaid, another sample of families who had fee-for-service Medicaid was selected and asked to enroll in the study but was not asked to agree to randomization (the second subsample). A self-selected HMO sample was chosen from the families with at least one member already in the study HMO plan at the time of the baseline interview.

Families that could not be located, either because their Medicaid records listed post office boxes rather than residential addresses or for other reasons, and families that spoke little or no English were excluded from the study. Any family belonging to a prepaid plan other than the study HMO was also ineligible.

Although families were informed that their enrollment in the study was voluntary, participation rates were high. Of the eligible sample, 80% (4433 of 5542 individuals) agreed to participate and 75% were actually enrolled. That is, only 268 people refused to participate after the initial informed consent. Families who did not enroll were not assigned to another group. There were no significant differences in average prior monthly Medicaid use between those whom the study located and those who were not located, between those who were eligible for enrollment and those who were ineligible, or between those who refused informed consent and those who accepted it. Neither did these groups differ significantly in terms of family size, average age, or percentage female, with the one exception that family size was larger in the eligible group (3.7) than in the ineligible group (3.4). Finally, the refusal rate was as low among random assignment families as it was among the self-select families.

Models

Three different dependent variables were examined, each one describing a sample child's use of health services during the 2-month period covered by the diary. These variables are whether the child (1) had a "regular" or checkup visit, (2) had a visit for acute care, and (3) visited an emergency room. There were up to three diaries for each child, each covering a different 2-month period. Analyses include a variance component correction to control for correlation among observations for a particular child. Logistic regressions were used throughout. Emergency room visits were modeled as conditional on having made any visit. The independent variables were selected based on previous studies of child health. Both dependent and independent variables are listed in Table 1.

In addition to plan status (in HMO or fee-for-service care), the explanatory variables include sex, race/ethnicity, age and age squared, and number of health problems. The distributions of most variables are unremarkable: there are a few more boys than girls, the average age of the children is 6, and relatively few existing health problems are reported (on average, about one condition in every two children).

About 13% of the children were described by their parents as White, 80% as Black, and about 5% as Hispanic. Because a large but unidentifiable number of Caribbeans (some of whom might have been classified as Hispanic) reported

TABLE 2—Determinants of Checkup and Acute Care Visits by Children Enrolled in a Medicaid HMO or in Medicaid Fee-for-Service Care: Odds Ratios and 95% Confidence Intervals (CIs)

	Type of Visit			
	Checkup[a]	Acute	Checkup[b]	Acute
	Odds Ratio (95% CI)	Odds Ratio (95% CI)	Odds Ratio (95% CI)	Odds Ratio (95% CI)
Enrolled in HMO plan	1.00 (0.72, 1.38)	0.72 (0.52, 1.01)	...	...
Enrolled in HMO plan, had no preexisting health problems			0.97 (0.69, 1.36)	0.65 (0.44, 0.95)
Enrolled in HMO plan, had at least one preexisting health problem	...	...	1.01 (0.74, 1.37)	1.08 (0.83, 1.39)
Randomly assigned	0.95 (0.67, 1.35)	0.86 (0.60, 1.23)	0.95 (0.67, 1.34)	0.93 (0.65, 1.33)
Female	0.90 (0.70, 1.14)	1.02 (0.79, 1.33)	0.90 (0.70, 1.14)	1.03 (0.79, 1.33)
White	1.39 (0.94, 2.05)	1.18 (0.73, 1.90)	1.39 (0.94, 2.05)	1.17 (0.73, 1.90)
Child's age in years	0.72 (0.63, 0.83)	0.69 (0.60, 0.79)	0.72 (0.63, 0.83)	0.69 (0.60, 0.79)
Child's age squared	1.02 (1.01, 1.03)	1.02 (1.01, 1.03)	1.02 (1.01, 1.02)	1.02 (1.01, 1.03)
Numbers of allergies listed for child at start of experiment	1.06 (0.80, 1.40)	1.36 (1.06, 1.75)	1.05 (0.73, 1.51)	1.20 (0.89, 1.62)
Number of serious health problems listed at start of experiment	3.02 (1.09, 8.36)	2.13 (0.63, 7.28)	3.03 (1.11, 8.30)	2.40 (0.84, 6.86)
Number of other health problems listed at start of experiment	1.19 (0.91, 1.56)	1.57 (1.21, 2.04)	1.18 (0.88, 1.58)	1.38 (1.09, 1.74)
Diary is from second wave	0.81 (0.55, 1.20)	0.58 (0.39, 0.88)	0.81 (0.55, 1.20)	0.56 (0.37, 0.83)
Diary is from third wave	0.64 (0.37, 1.11)	0.50 (0.27, 0.92)	0.64 (0.37, 1.12)	0.49 (0.27, 0.91)
Diary describes April to September	1.28 (0.88, 1.85)	1.87 (1.26, 2.78)	1.27 (0.88, 1.85)	1.83 (1.23, 2.73)

[a]Columns 1 and 2 contrast all HMO enrollees with all fee-for-service enrollees.
[b]Columns 3 and 4 contrast HMO enrollees with chronic conditions and HMO enrollees without chronic conditions with fee-for-service enrollees.

TABLE 3—Determinants of Emergency Room Visits by Children Who Used HMO or Fee-for-Service Medical Care: Odds Ratios and 95% Confidence Intervals (CIs)

	Odds Ratio	95% CI
Enrolled in HMO plan	1.06	0.62, 1.80
Randomly assigned	1.17	0.65, 2.11
Female	1.07	0.69, 1.65
White	1.15	0.52, 2.58
Child's age in years	0.80	0.64, 1.01
Child's age squared	1.02	1.00, 1.04
Total number of health problems listed as start of experiment	1.13	0.92, 1.39
Diary is from second wave	0.50	0.25, 0.99
Diary is from third wave	0.39	0.13, 1.13
Diary describes April to September	1.82	1.04, 3.20

themselves as Black, only White vs non-White children are contrasted in the models.

Analyses controlled for whether the child was randomly assigned or self-selected to his or her plan status (about equal numbers of diaries are from each group), whether the diary describes months from fall and winter (October to March) or from spring and summer (April to September), and which wave of data collection (first, second, or third) the diary is from. The self-select and random assignment groups were quite similar to each other in age, sex, ethnic composition, and rates of outpatient visits.

Results

Odds ratios from logistic regression models are shown in Tables 2 and 3. Table 2 presents the simplest model of the determinants of checkup and acute care visits. Initial tests showed no significant differences in use between the self-select and random assignment groups within each plan status, as might be expected since participation rates were similar. Because the use of care within the HMO did not differ between those who self-selected the HMO and those who were assigned to the HMO, a single indicator was used for the two HMO groups. Similarly, there were no significant differences between the two fee-for-service groups, which were also combined into a single variable. Table 2 shows that children enrolled in an HMO have a lower probability of an acute care visit. Although the coefficient on HMO enrollment just fails to meet the standard 5% test for statistical significance ($P = .056$), it is large enough to indicate that HMO enrollment triggers a real and important reduction in acute care visits. (In fact, the coefficient on HMO enrollment for acute care visits was negative and significant [$P = .020$] before we corrected the regression standard errors for covariance between repeated observations from the same child and from the same household. With this covariance correction, the coefficient estimate did not change but the estimated standard error became larger; consequently, the statistical significance just dipped below the 5% threshold [$P = .056$].) The predicted rate of acute care visits for plan children was only 14% while for fee-for-service children it was 23% (controlling for all other factors in the model). In contrast, on checkup visits, plan members and fee-for-service children were indistinguishable. Although possible reasons for these differences are discussed in detail in the Discussion section below, three obvious possibilities are noted here: (1) the HMO may provide better, more complete care at the time of the checkup, so fewer acute care visits are necessary; (2) the HMO may reduce unnecessary use among basically healthy children but may serve children with serious or frequent health problems at an

appropriately higher level; or (3) the HMO may respond to the incentives inherent in any capitation scheme and may underprovide needed services (e.g., by making it difficult to come in for acute care). This third explanation is not supported by the data in Table 2, however. If the HMO were withholding services, presumably the effect should also appear in rates of checkup visits, which it does not.

Child age has the expected effect on medical visits. The probability of any type of visit falls at a decreasing rate with age, bottoming out around age 10. Although boys typically are frailer than girls, sex is not a significant predictor of either acute or checkup visits when prior health status and age are controlled.

Children's prior health problems have different effects on acute care and checkup visits. A child with a major disabling condition (blindness, diabetes, cerebral palsy, mental retardation, cancer, missing limbs) was significantly more likely than a nondisabled child to visit a health provider for routine care (probability of a checkup = 41%), while a child with an allergy (eczema, asthma, hay fever, other skin and respiratory allergies) or "other" (nondisabling) condition had an above-average likelihood of an acute care visit.

Thirty-one percent of Whites but only 19% of non-Whites had a recent checkup visit. Although this variable is significant at only the 10% level, in the model reported here it had a significance level of .046 before correction for correlated observations.

The final variables that were considered are aspects of the survey administration: method of assignment and timing of the diary. The random assignment groups did not differ significantly from the self-selected groups in terms of use. The season of the year, however, did affect use. This site is in Florida, where the summer months bring significantly more acute care problems, presumably for ailments associated with exposure to sun and humidity and with moving between air conditioning and extreme heat. The coefficients on wave 2 and wave 3 show respondents reporting fewer acute care visits in later waves of the survey, independently of the season. This pattern, which cannot be readily explained, is consistent with the usage patterns observed in the claims data.

Table 2 indicates that the structure or practices of the HMO led to fewer acute care visits among enrolled children. Several hypotheses were tested that might explain this effect. The first was that the HMO might succeed in reducing care among children with few needs but might still see children with chronic problems at the same rate as under fee-for-service care. The fourth column in Table 2 supports this theory. The decline in acute care visits is concentrated among children with no health problems identified at the start of the experiment; their probability of an acute care visit in the 2-month period covered by a diary is only 11.5%, compared with a 16.7% probability for equally healthy children in the fee-for-service system. Children with one or more problems at the outset of the study had equal acute care visit probabilities in the two systems. As the third column shows, the likelihood of a checkup visit was not affected by plan enrollment for either group.

The hypothesis that physicians might provide more comprehensive care during visits to the HMO was tested using data on the content of acute and well care visits (e.g., whether the child received a urine test, throat culture, etc.). However, t tests revealed no significant differences between visit content in the HMO and the fee-for-service systems. Although children in the plan were somewhat more likely to have a urine test during a preventive visit ($P = .07$), they were somewhat less likely to receive a throat culture during an acute care visit ($P = .09$). Five other comparisons showed no significant differences.

Finally, the study sought to determine if the HMO selectively reduced emergency room visits relative to fee-for-service use. Overall, 2.4% of children used the emergency room for a checkup in a 2-month period and 4.8% visited it for an acute problem. Viewed differently, nearly one quarter of all acute visits and 9% of all checkup visits occurred in the emergency room. To determine whether, among those who used medical services, HMO enrollees were less likely to use the emergency room, the model made the probability of any visit to the emergency room conditional on the child having had at least one checkup or acute visit. As Table 3 shows, there are no significant differences between HMO and fee-for-service enrollees in their use of the emergency room. More elaborate models, examining acute and checkup emergency room visits separately, also failed to reveal any significant determinants of use.

Discussion

This paper has shown that an HMO reduced acute care visits for children on Medicaid but did not reduce emergency room use. The deterrent effect of HMO membership on acute care visits is not found across the board for all children. On the contrary, it is concentrated in a particular group: children with no health problems at the start of the experiment. Thus, the HMO appears to be targeting its services to children with the greatest health care needs rather than curtailing services across the board.

The question remains: Why should healthy HMO plan members have fewer acute visits than fee-for-service children do? Earlier we suggested three possible reasons: the plan provides more complete care during each preventive visit, it reduces necessary acute care, or it reduces unnecessary visits. Data on the content of well care visits do not offer any support for the first hypothesis. Nor do they support the second explanation: that the HMO is deliberately making care inaccessible to patients. If this were the case, presumably we would observe a negative effect of plan enrollment on checkup visits. However, the reduction in visits is limited to acute care visits by the most healthy children, which suggests that the HMO is able to "rationalize" its delivery of services rather than to merely ration services across the board. The assignment of a case manager to the HMO patients may have helped them receive advice over the telephone without making a visit. Since we have no data on telephone consultations, we cannot confirm this hypothesis. In contrast to the findings in Ware et al.,[13] patients with preexisting health conditions apparently received as much care in the HMO as they did in the fee-for-service system.

Although visits to the emergency room accounted for nearly one quarter of all the Medicaid children's acute care visits, the HMO did not selectively reduce emergency room use. In contrast to most previous studies, we found that the proportion of all HMO pediatric visits that occurred in the emergency room did not differ significantly from the proportion for fee-for-service patients. This may be because the HMO was sponsored by and located in the hospital with a large emergency department serving Medicaid patients. Thus, as was the case in Hurley et al.,[9] this plan may have had little incentive to divert patients from emergency room use.

The finding that the HMO reduced overall use by targeting health care services to those with the greatest health needs is encouraging news for advocates of Medicaid HMOs since it suggests that, in this experiment, the HMO may have

reduced unnecessary use without reducing care for children who need it. However, we must conclude on a cautious note. First, our results are based on practice at a single HMO. Second, while we have no evidence that the HMO was not providing needed services, neither can we claim that the HMO was remedying all deficits in care in a population vulnerable to inadequate care. □

Acknowledgments

This research was supported by the Health Care Financing Administration, US Department of Health and Human Services, Cooperative Agreement 99-C-98489/9-05. The prepaid managed health care demonstration was funded by the Robert Wood Johnson Foundation.

The authors thank Jerene Kelly for her help in preparing the manuscript.

References

1. Aday LA, Fleming G, Anderson R. *Access to Medical Care in the U.S.: Who Has It, Who Doesn't.* Chicago, Ill: Pluribus Press and Center for Health Administration Studies; 1984.
2. Miller M. The role of substitutes in policy analysis: acute care services in state Medicaid programs. *J Health Polit Policy Law.* 1988;13:499–524.
3. Long SH, Settle RF, Stuart BC. Reimbursement and access to physicians' services under Medicaid. *J Health Econ.* 1986;5:235–251.
4. Cohen JW. Medicaid policy and the substitution of hospital outpatient care for physician care. *Health Serv Res.* 1989;24:33–66.
5. Kasper JD. The importance of type of usual source of care for children's physician access and expenditures. *Med Care.* 1987;25:386–398.
6. Hurley RE, Freund DA, Taylor DE. Emergency room use and primary care case management: evidence from four Medicaid demonstration programs. *Am J Public Health.* 1989;79:843–847.
7. Bonham GS, Barber GM. Use of health care before and during Citicare. *Med Care.* 1987;25:111–119.
8. Davidson S, Fleming G, Hohlen M, Manheim L, Shapiro B, Werner S. *Physician Reimbursement and Continuing Care under Medicaid: A Demonstration (The Children's Medicaid Program).* Elk Grove Village, Ill: American Academy of Pediatrics; May 1988.
9. Hurley RE, Freund DA, Taylor DE. Gatekeeping the emergency department: impact of a Medicaid primary care case management program. *Health Care Manage Rev.* 1989;14:63–71.
10. Manning WG, Leibowitz A, Goldberg GA, Rogers WJ, Newhouse JP. A controlled trial of the effect of a prepaid group practice on utilization. *N Engl J Med.* 1984;310:1505–1510.
11. Perkoff GT, Kahn L, Haas PJ. The effects of an experimental prepaid group practice on medical care utilization and cost. *Med Care.* 1976;14:432–449.
12. Luft HS. *Health Maintenance Organizations: Dimensions of Performance.* New York, NY: John Wiley & Sons Inc; 1981.
13. Ware JE Jr, Rogers WH, Davies AR, et al. Comparison of health outcomes at a health maintenance organization with those of fee-for-service care. *Lancet.* 1986;1:1017–1022.
14. Valdez RB, Ware JE Jr, Manning WG, et al. Prepaid group practice effects on the utilization of medical services and health outcomes for children: results from a controlled trial. *Pediatrics.* 1989;83:168–180.
15. Hohlen MM, Manheim LM, Fleming GV, et al. Access to office-based physicians under capitation reimbursement and Medicaid case management: findings from the children's Medicaid programs. *Med Care.* 1990;28:59–68.

The Role of State Health Departments in Formulating Policy: A Survey on the Case of AIDS

Leonard Robins, PhD, and Charles Backstrom, PhD

ABSTRACT

Objectives. Public health departments have been criticized for not being on the cutting edge of health policy formation in the United States and, in particular, for being ineffective in dealing with acquired immunodeficiency syndrome (AIDS). This study examines the attitudes and influence of public health officials on this problem.

Methods. A survey of state health department officials was supplemented by surveys of state legislature health committee chairs and hospital association directors, as well as by interviews with key AIDS policymakers and observers in six states.

Results. Both in their own eyes and in the eyes of legislative leaders and hospital association officials, public health officials were considered to be the principal initiators of public policy on AIDS.

Conclusions. Contrary to the Institute of Medicine's 1988 report, state health departments are leaders in the fight against AIDS at the state level. Moreover, we argue that public health departments should be judged on how effectively they perform functions for which they have primary legal responsibilities, not for other roles for which they have been given neither responsibility nor resources. (*Am J Public Health.* 1994;84: 905–909)

Introduction

In 1986 the Institute of Medicine established a Committee on the Future of Public Health to address a growing perception that the nation has "allowed the system of public health activities to fall into disarray."[1] After 2 years of study, the committee issued its report and concluded that "unfortunately, the findings of this committee confirm the concerns that led to the study."[1]

The committee report chose acquired immunodeficiency syndrome (AIDS) as an example of how poorly the public health system works. It emphasized the danger of public health officials having overly broad statutory power for imposing quarantine and implied that health agencies overemphasize surveillance at the expense of education. Finally, it asserted that health agencies have not supplied leadership in dealing with this new disease.[1]

The report gave several reasons for the general inadequacy of public health departments. Among the most important was its belief that public health officials were both naive about and contemptuous of the political process and hence lacking in political power.[1]

The purpose of this article is to examine these conclusions. But instead of focusing on the influence of public health professionals on the overall health care system of the United States, we ask a different question: How well are these professionals performing the classic public health functions that are still within their jurisdiction? We propose to determine this by emphasizing their role in the fight against AIDS—the greatest communicable disease challenge of our time and an issue that obviously offers an important role for public health agencies to play.

Methods

As part of an ongoing study of AIDS politics,[2] in the spring and fall of 1989, questionnaires concerning their state's policies on AIDS were sent to the nation's 50 state health departments, the 76 areawide hospital associations in the United States, and the chairs of all state House and Senate health committees. Those not responding within 3 weeks were sent a follow-up letter with another questionnaire. Ultimately, usable responses were received from 42 (84%) of the state health departments; 51 of the 76 (67%) state and regional hospital associations representing 39 (78%) of the states; and 68 legislative committees in 43 (86%) of the states. A number of the questions were common to all sets of respondents; other questions were geared to the special concerns of each category of respondents. Although the response rate was gratifyingly high, the total number of responses was inevitably relatively small. Because the universe of states was surveyed, no tests of statistical significance were used.

In addition to the surveys, field interviews were conducted with AIDS policymakers in six states—Arizona, California, Georgia, Illinois, New York, and Minnesota—in the fall and winter of 1989 to 1990.

This paper mainly reports the results from the state health departments survey, supplemented by selected findings from the other surveys and by occasional insights from the interviews.

Results

The Statutory Authority of Public Health Agencies

The Institute of Medicine's committee report asserted that "the statutory base of public health is poorly suited to dealing with AIDS." In particular, the committee was concerned about overly

Leonard Robins is with the Public Administration Program, Roosevelt University, Chicago, Ill. Charles Backstrom is with the Department of Political Science, University of Minnesota/Twin Cities, Minneapolis.

Requests for reprints should be sent to Charles Backstrom, PhD, Department of Political Science, University of Minnesota, 1414 Social Sciences Bldg, 267 19th Ave S, Minneapolis, MN 55414.

TABLE 1—Priorities Given Various AIDS Efforts by State Health Departments (n = 40)

	Priority, %			
	Highest	High	Low	Should Not Do
Getting high-risk persons to change sexual behavior	70.0	27.5	2.5	0.0
Reaching minorities on AIDS prevention	52.5	42.5	5.0	0.0
Getting high-risk persons to test for HIV	43.9	51.2	4.9	0.0
Informing public about how AIDS is spread	34.1	56.1	9.8	0.0
Knowing extent of AIDS problem	26.8	73.2	0.0	0.0
Providing sufficient drug rehabilitation	32.5	47.5	20.0	0.0
Restricting persons with HIV from spreading it	24.3	51.4	13.5	10.8
Getting heterosexuals to practice safe sex	14.6	65.9	17.1	2.4

restrictive powers, including full isolation and quarantine, that had earlier been given to health departments.[1]

The relevant issue is whether state health departments endorse a policy of quarantining persons who have AIDS or who test positive for human immunodeficiency virus (HIV), or if they instead support a less restrictive policy. One alternative would be reliance on civil commitment procedures, to be used if a person who is HIV positive persists in behavior that endangers others even after being warned to cease. This is similar to the manner in which someone who is mentally ill and thought to need control is handled. Unlike quarantine, civil commitment is hedged with due process protections. Another option is using existing criminal statutes, under which a charge of assault could be brought against someone who is positive for HIV and who wantonly or carelessly infects others.

Respondents to our survey were asked whether their present state policy included quarantine, especially for AIDS; civil commitment authority; and/or the criminalization of risky behavior. They were also asked whether the health department favored such policies. In only one instance was quarantine for AIDS the theoretically official policy of the state, and in only that instance was quarantine favored. Thus, it is clear that quarantine is overwhelmingly opposed by state health officials and that, whether taking the lead or acting in alliance with others, they have successfully prevented it from becoming policy.

As for civil commitment, just over half of all respondents said they favored it, just over one quarter opposed it, and 43% said their states actually had such a policy. On the other hand, 38% of the states had criminal liability as their policy, which was favored by only one third of the health departments and opposed by the same proportion, with the rest uncertain or not responding.

In Minnesota,[3] interviews with health and legal officials dealing with AIDS revealed that state health officers were fully aware of the danger of untrammeled governmental powers. General quarantine statutes from the earliest days of statehood gave unrestricted power to the chief state health officer to isolate anyone with a serious communicable disease who appeared to be a threat to the public health. However, the state health department actually feared that the use of this statute—which contained no due process provisions such as notice and hearing—against AIDS would open them to justifiable complaint. As a result, they sought and obtained from the legislature a modification to make their powers conform to modern practice in civil commitment of the mentally ill. They did not want the quarantine statute repealed, however, because it provided them with a cover for opposing other restrictive laws about AIDS that might well have been passed. In fact, as evidence that the health department was actively protecting the larger public, they pointed to two, but only two, instances (a female and a male prostitute who were infected) in which they used the authority to restrain persons who did not respond to orders to desist.

It would appear that the state health officers are heeding the committee report's warning that public health statutes give state health departments overly broad and restrictive powers. Not only do they not want additional draconian powers, but they in fact oppose use of "solutions" such as quarantine. Contrary to the common assumption that any bureaucracy always desires more power, they show a sophisticated understanding of what constitutes appropriate legal powers against AIDS.

Data Gathering and Analysis

The Institute of Medicine's report noted that "some health agencies are accused of overemphasizing surveillance at the expense of preventive efforts such as education."[1] Given the use of the word *some*, this is undoubtedly true. Surveillance is the traditional core of classic public health, and the inevitable interface of AIDS education with sex and drug use ensures that such education will be controversial. Nonetheless, as Table 1 shows, in terms of efforts to deal with the disease, state health departments are far from being totally preoccupied by surveillance issues.

Although in no state did public health officials assign a low priority to knowing the extent of AIDS in their state, five other AIDS efforts were considered by a greater proportion of state health departments to have highest priority. Surveillance, however, was ranked as highest priority by more health officials than were two other listed efforts. Thus, while all state health departments consider surveillance important, many are equally or more strongly committed to more activist aspects of AIDS control.

Influence on Policy

The institute report makes the following charge:

> Among the many groups and individuals, public and private, engaged in fighting AIDS, health agencies have not taken a clear initiative in supplying leadership, and the public is unclear about what level of government it should look to for guidance or what it can appropriately and realistically expect any particular health agency to do.[1]

It is indeed undoubtedly true that the public is unsure of which level of government is responsible for various activities. Alas, that is true of almost all public policy problems, save perhaps for national defense at one extreme and fire protection at the other. The only concern of the public is whether anyone is doing the job. In this case, the issue is whether state health departments have a leading role in AIDS policy development.

As evidence that health departments are often not critical players in the

TABLE 2—Relative Assessment by Health Departments (n = 42), Legislative Chairs (n = 68), and Hospital Association Directors (n = 47) of the Influence of Various Actors on State AIDS Policymaking

	Very Influential, %			Some Influence, %			Little Influence, %			No Influence, %		
Policy Actor	Health Department	Legislative Chair	Hospital Association	Health Department	Legislative Chair	Hospital Association	Health Department	Legislative Chair	Hospital Association	Health Department	Legislative Chair	Hospital Association
State department of health administrators	59.5	67.6	72.3	38.1	30.9	25.5	2.4	1.5	2.1	0.0	0.0	0.0
Elected officials												
Legislative leaders	57.1	48.5	46.8	35.7	45.6	44.7	7.1	5.9	6.4	0.0	0.0	2.1
Governor	42.9	20.6	42.6	40.5	52.9	29.8	9.5	22.1	25.5	7.1	4.4	2.1
Health providers												
State medical association	23.8	29.9	19.1	61.9	62.7	40.4	14.3	7.5	2.1	0.0	0.0	38.3
State hospital association	14.3	25.0	23.4	64.3	61.8	46.8	16.7	11.8	25.5	2.4	1.5	4.3
Insurance industry	7.1	11.8	17.0	52.4	57.4	55.3	31.0	27.9	23.4	7.1	2.9	4.3
Affected groups												
Persons living with AIDS	9.5	15.2	6.4	40.5	43.9	34.0	42.9	39.4	51.1	7.1	1.5	8.5
Organized gay groups	9.5	10.3	6.4	54.8	29.4	34.0	26.2	44.1	42.6	7.1	16.2	17.0
Ideological organizations												
Conservative ideological groups	14.3	2.9	4.3	28.6	26.5	38.3	38.1	50.0	48.9	16.7	20.6	8.5
Liberal ideological groups	2.4	4.4	2.1	38.1	27.9	31.9	42.9	54.4	53.2	14.3	13.2	12.8
Mainline church leaders	4.8	1.5	4.3	47.6	35.3	34.8	42.9	50.0	50.0	4.8	13.2	10.9
Fundamentalist church leaders	2.4	0.0	6.5	35.7	20.6	23.9	42.9	50.0	45.7	19.0	29.4	23.9

Note. Percentages of a single rating group may not total 100.0 across because the answers of those who volunteered "Don't know" are not tabulated; percentages may exceed or not equal 100.0 because of rounding.

formulation of AIDS policies, the report quotes a legislative staffer:

> The legislature has been the leader. . . . Such leadership should have come from the Department of Health Services. The state health director knows less than I do about what's happening in the state![1]

Our study finds otherwise.

Table 2 presents the relative assessment by state health department heads, legislative health committee chairs, and hospital association directors of the influence of various actors on state AIDS policy. Potential actors are grouped into the following categories: health department heads, elected officials, health providers, affected groups, and ideological groups.

Perhaps not surprisingly, health departments rated themselves as being the most influential actors on AIDS policymaking in their states. Sixty percent felt they were very influential and 38% thought they were somewhat influential in setting AIDS policy; only one respondent believed they had little influence and not a single one asserted they had no influence.

Strong confirmation of the influential role that health departments play in AIDS policymaking is provided by the other groups surveyed. Legislative health committee chairs and hospital association professionals ranked the health departments clearly higher in influence than they ranked any other actor.

The next most influential AIDS policymakers were seen to be legislative leaders. Well over half (57%) of the health departments perceived legislative leaders as very influential and more than one third (36%) said somewhat influential. This assessment was seconded by the legislative chairs and the hospital association professionals.

Although somewhat less important, the governor was the only other actor believed to have substantial influence on AIDS policymaking. Forty-three percent of health departments rated their governor as being very influential and 41% said somewhat influential. Hospital associations ranked the governor similarly. Legislative respondents, on the other hand, ranked the governor substantially lower.

At a distinctly lower level of influence stand health provider associations. Only about a quarter of the health departments rated their state medical associations as highly influential, but more than three in five classified them as

TABLE 3—Perceived Importance of Various Functions of State Health Department Regarding AIDS Policy (n = 42)

Functions	Highest Importance, %	High Importance, %	Important, %	Not Important, %
Devising appropriate statewide policies	54.8	42.9	2.4	0.0
Making the public more aware of AIDS	45.2	38.1	16.7	0.0
Leadership of health professionals	40.5	45.2	11.9	2.4
Professional adviser to the legislature	31.0	50.0	14.3	4.8
Professional adviser to the governor	29.3	48.8	14.6	7.3
Developing innovations in agencies' programs	28.6	42.9	28.6	0.0

Note. Percentages may exceed 100.0 because of rounding.

TABLE 4—Respondents' Ratings of the Effectiveness of Possible "Liberal" and "Conservative" Strategies for Adopting AIDS Policies (n = 39)

	Effectiveness of Strategy, %			
	Good	Uncertain	Poor	Edge: Good-Poor
"Liberal" strategies				
Emphasize human interest; people suffer	70.0	22.5	7.5	62.5
Argue the effectiveness of liberal measures	64.1	28.2	7.7	56.4
Stress community values; helping others	43.6	41.0	15.4	28.2
Urge keeping AIDS out of electoral politics	38.5	48.7	12.8	25.7
Encourage threatened groups to work together	30.8	43.6	25.6	5.2
Make stand on protection of civil liberties	33.3	30.8	35.9	−2.6
"Conservative" strategies				
Point out the need to protect innocent people	87.2	10.3	2.6	70.7
Make stand on the duty to protect public	64.1	30.8	5.1	59.0
Dramatize the threat to the general public	61.5	33.3	5.1	56.4
Emphasize how those at risk differ from majority	43.6	33.3	23.1	20.5
Encourage greater cohesion among conservatives	33.2	56.4	10.3	22.9
Argue that AIDS is a valid election issue	17.9	59.0	23.1	−5.2

having some influence. Legislative chairs pegged the medical associations somewhat higher, but hospital association directors ranked them far lower in influence. As for hospital associations, 14% of the health departments believed them to be very influential in AIDS policymaking, but 64% said they had only some influence. Legislative chairs ranked them higher, but only 25% of these respondents classified hospital associations as very influential, and slightly fewer of the hospital associations (23%) ranked themselves as being very influential.

Weakest among provider groups is the insurance industry, which only 7% of the health departments rated as very influential. Legislative chairs and hospital associations ranked them somewhat higher but still below the other providers.

Affected groups such as persons living with AIDS and organized gays were seen by fewer than 1 in 10 health departments as being very influential. At an even lower level of influence were ideological organizations and church leaders. Only conservative organizations were seen as highly influential by any rating group, being so classified by 14% of the health departments.

Political scientists have long debated the usefulness of reputation as a surrogate measure for actual public policy influence since there may well be a low correlation between who is thought to have influence and who has it in reality. This doubt is especially persuasive when those whose opinions are asked are not experts on the subject and are not themselves involved in decision making on the matter at hand. In this case, however, those responding were themselves specialized elites working in the same subject matter area: AIDS.

Understanding the Political Process

The Institute of Medicine report was also very critical of public health officials for their lack of political skill and understanding:

> Too frequently during its investigations, the committee heard legislators and members of the general public castigate public health professionals as paper-shufflers, out of touch with reality, and caught up in red tape.... Yet many public health professionals who talked with us seemed to regard politics as a contaminant of ideally rational decision-making processes rather than as an essential element of democratic governance. We saw much evidence of isolation and little evidence of constituency-building, citizen participation, or continuing (as opposed to crisis-driven) communication with elected officials or with the community at large.[1]

Although it is undoubtedly true that some public health officials lack political sophistication, our study results contradict the view that public health officials are generally politically inept and ignorant, at least with regard to AIDS. Two questions from our survey included political content most pertinent to this subject.

First, to test the view that state health departments are preoccupied with their internal procedures, respondents were asked how important they regarded each of various aspects of their job as it concerned AIDS (Table 3). Developing innovations in the agency's AIDS program—an inward-focused activity—was found to be the least important function in the eyes of the respondents. Although all respondents believed that internal innovations are important, fewer than one third (29%) considered it the highest priority, the lowest among the six categories suggested.

It should be acknowledged that higher importance given by state health departments to external rather than to internal functions does not automatically mean that these departments are doing a good job of constituency building or fostering citizen participation. It does strongly suggest, however, that they are not preoccupied with paperwork and that communication with elected officials is more than episodic. Evidence earlier presented ranking their influence on AIDS policy reinforces this view.

Second, to investigate the charge that public health professionals are ignorant of and have contempt for the political

process, respondents were asked to rate the effectiveness of a wide array of various "liberal" and "conservative" strategies for influencing AIDS policy. We then computed the "edge" by which respondents label a strategy as good over those who think it is poor (Table 4). Least support among liberal strategies (receiving 33% and the only negative edge) was given to stressing civil liberties as an effective approach. Conversely, an overwhelming majority (70% with a 63% edge) considered emphasizing human suffering as a good strategy from the liberal perspective. A massive majority (87% with a 71% edge) endorsed emphasizing the need to protect innocent victims as a good conservative strategy. On balance, a small group (23% with a −5% edge) thought that arguing that AIDS is a valid election issue would be a poor strategy.

To us as political scientists, these judgments show political sagacity. While the analysis of these responses as indicators of political sophistication must inevitably be subjective, interviews with non-health policy actors and observers of the legislative process in the six states visited generally show agreement with the health departments' assessments.

Discussion

Our study of AIDS policymaking found that, in dealing with AIDS, public health departments are very influential. Moreover, both our surveys and our interviews convinced us that public health professionals promoted their ideas with a clear understanding of their political environments and the strategies appropriate in them. These findings dramatically differ from the Institute of Medicine's indictment of public health professionals, in which public health was characterized as being in disarray and politically amateurish.

How do we account for these different perspectives? One explanation may be that the negative view expressed in the institute's report is generally correct and that our findings refute its assertions only when applied to AIDS. But the context in which public health operates has also changed with the challenge of AIDS. A massive communicable disease threat at a time when antibiotics had seemingly downgraded the necessity for classic public health presented the opportunity for public health agencies once again to move to center stage instead of being relegated to unglamorous activities like restaurant inspection.

More fundamentally, we suggest that our conclusions arise from different evaluation criteria. The Institute of Medicine spent a good deal of time noting that many important health issues are not under the jurisdiction of public health departments. This, however, is a strange criterion. It is perfectly appropriate to recommend the consolidation of health functions under a superagency, but it is wrong to use the failure of this to occur as evidence of the inadequacies of public health professionals. Why should we be surprised that environmentalists or mental health specialists want their own agencies, or that they have been able to succeed in turning those agencies into separate departments, since public health professionals themselves do not like to be a division of a superagency?

Similarly, the criticism that public health agencies have not acted to institute major reforms of the health delivery system is unfair. Unfortunately, the US government has not yet enacted national health insurance, but this requires a political decision at a more fundamental level than bureaucrats can provide.

Thus, we show that these maligned public health agencies reacted with professional competence and political skill (two not unrelated aspects) to deal with the threat of AIDS. Perhaps public health professionals are rarely thought about in routine times, but they are nonetheless quite competent when a public health emergency arises in which they are needed. Their apparent decline is true at most only in a relative sense. □

Acknowledgments

An earlier version of this article was presented before the American Political Science Association on September 5, 1992, in Chicago, Ill.

We wish to thank Theodor J. Litman, PhD, and three anonymous reviewers for helpful comments on drafts of this work.

References

1. Institute of Medicine. *The Future of Public Health.* Washington, DC: National Academy Press; 1988.
2. Robins L, Backstrom C. The politics of AIDS. In: Litman T, Robins L, eds. *Health Politics and Policy.* 2d ed. Albany, NY: Delmar Publishers; 1991:370–386.
3. Backstrom C, Robins L. *The Minnesota Response to AIDS.* Minneapolis, Minn: University of Minnesota, Center for Urban and Regional Affairs; 1992.

The Use of Nonoxynol-9 for Protection against Cervical Gonorrhea

Sharon S. Weir, MA, MPH, Paul J. Feldblum, PhD, Leopold Zekeng, PhD, and Ronald E. Roddy, PA-C, MPH

ABSTRACT

Objectives. Although condoms are the best defense against sexually transmitted disease, little is known about the effectiveness of female-controlled methods containing nonoxynol-9 as backup protection when condoms are not being used.

Methods. To assess the extent to which nonoxynol-9 protects women against gonorrhea, a cohort of 303 female sex workers (prostitutes) in Yaounde, Cameroon, were asked to use condoms and suppositories containing nonoxynol-9 at every sexual encounter and to record daily sexual activity and use of condoms and suppositories on coital logs that were reviewed monthly. Evidence of gonorrheal infection was based on a positive gonorrhea culture. Stratified analysis and proportional hazards regression were used to estimate rate ratios.

Results. Forty-one women enrolled in the study were excluded from the current analysis. The estimated incidence of gonorrhea was 6.2 infections per 100 person-months of observation. Incidence rate ratios estimated from proportional hazards regression models controlling for condom use showed that using nonoxynol-9 during acts not protected by condoms reduced the risk of infection.

Conclusions. Although the protective effect of condoms against sexually transmitted disease is greater than that afforded by nonoxynol-9, using nonoxynol-9 when condoms are not used is a far better strategy in gonorrhea prevention than using no method at all. (*Am J Public Health.* 1994;84:910–914)

Introduction

When used properly, condoms are effective in preventing most sexually transmitted diseases, including cervical gonorrhea.[1-5] However, the underuse of condoms among people at high risk for sexually transmitted diseases suggests that there is an urgent need for alternative methods.[3,6,7] Clinical and epidemiological studies indicate that the spermicidal compound nonoxynol-9 can be effective in vitro[8,9] and in vivo[10-15] in preventing gonorrheal infection. Spermicides have some desirable properties for women not shared by condoms. Most notable, a woman can insert a spermicidal compound ahead of time, without the need for her partner's direct involvement or cooperation.

Between 1989 and 1990, a cohort study of female sex workers (prostitutes) in Yaounde, Cameroon, was conducted to assess the protective effect of nonoxynol-9 against human immunodeficiency virus (HIV) infection. Study results regarding the association between use of this compound and HIV infection have been published elsewhere.[16] This paper presents a further analysis of those cohort data to assess the extent to which nonoxynol-9 afforded protection against cervical gonorrhea.

Methods

Public health authorities enrolled 303 Cameroonian female sex workers in the study. Eligibility criteria included (1) age of at least 18 years; (2) no pregnancy during the previous 42 days (by self-report); (3) no history of adverse reaction to a spermicidal product; (4) negative enzyme-linked immunosorbent assay (ELISA) HIV antibody test; and (5) willingness to use the study products, maintain a coital log (a calendar on which coitus, condom use, and spermicide use could be marked daily), and return monthly for 1 year for follow-up visits.

At admission, following an explanation of the study and their written informed consent, participants were administered a brief questionnaire to obtain information on their sociodemographic characteristics, sexually transmitted disease symptoms, and sexual practices. The women were advised to use condoms and vaginal suppositories containing nonoxynol-9 at every intercourse, and they were instructed in proper use of these products. Plain lubricated (nonspermicidal) latex condoms and vaginal suppositories containing 100 mg of nonoxynol-9 were provided at admission and were freely available throughout the study. (The other ingredients in the suppositories were polyethylene glycol, benzethonium chloride, citric acid, D&C red no. 21, D&C red no. 33, methylparaben, and water.) The women were instructed to return monthly, at which time they were interviewed about their sexual practices, use of medication, and sexually transmit-

Sharon S. Weir, Paul J. Feldblum, and Ronald E. Roddy are with Family Health International, Research Triangle Park, NC. Leopold Zekeng is with the Ministry of Health, Yaounde, Cameroon.

Requests for reprints should be sent to Sharon Weir, MA, MPH, c/o Tom Grey, Family Health International, PO Box 13950, Research Triangle Park, NC 27709.

This paper was accepted January 31, 1994.

Note. The views expressed in this article do not necessarily reflect those of the funding agency.

Editor's Note. See related editorial by Potts (p 890) and letter by Feldblum and Weir (p 1032).

ted disease symptoms during the past month; were given a pelvic examination with gonorrhea culture; and had their coital logs reviewed and tallied by project staff who did not know whether the women were infected with gonorrhea. (The coital logs contained a picture of a condom and a suppository for each act. The woman marked the method used, if any, with an X. The logs were completed each day.)

All participants were offered a small monetary incentive for transportation costs (equivalent to the cost of a shared taxi ride) at each visit; free condoms and vaginal nonoxynol-9 suppositories throughout the study; free physical examinations, diagnostic tests, and appropriate sexually transmitted disease therapy for any infections diagnosed during the study; and pre- and posttest counseling by study staff. The protocol was approved by the Cameroon National AIDS Control Service and the Family Health International Protection of Human Subjects Committee.

Classification of Exposure

The classification of exposure to nonoxynol-9 for each participant was based on coital log data, taking into account condom use. First, the cumulative number of coital acts for each participant was determined by adding together the number of acts reported on monthly coital logs until one of the following occurred: incident gonorrheal infection, missing coital log data or missing gonorrhea test result data, other loss to follow-up, or end of the 12-month study. Next, each coital act was classified into one of four categories representing the four possible combinations of condom and nonoxynol-9 suppository use (i.e., concurrent condom and suppository use; only suppository use; only condom use; and neither condom nor suppository use). In initial life table and proportional hazards regression analyses, acts protected by condoms alone and acts protected by condoms and nonoxynol-9 together were combined into the category of acts protected by condoms because the initial study question was whether nonoxynol-9 afforded protection during acts when condoms were *not* used. In subsequent exploratory analyses, all four categories were used.

Classification of Disease

Evidence of gonorrheal infection was based on a positive gonorrhea culture (typical colony morphology, oxidase positive, and gram stain positive). Swabs were taken from the endocervix, plated directly onto modified Thayer-Martin media, and incubated under carbon dioxide conditions. Plates were read at 24 hours and, if no growth was apparent, at 48 hours.

TABLE 1—Mean Percentage of Coital Acts Protected by Condoms and Nonoxynol-9 Suppositories Together, Condoms Only, Nonoxynol-9 Suppositories Only, or Neither Method

	Mean % of Acts Protected	SD	Range, %
Both methods	40.5	22.8	0–100
Condoms only	22.2	15.5	0–98
Nonoxynol-9 suppositories only	24.9	16.2	0–82
Neither method	12.4	13.5	0–83

Statistical Methods

Crude assessments of the protective effect of condom use and of nonoxynol-9 use during acts not protected by condoms were made by graphical representation of the life table curves of time until incident gonorrhea for varying levels of use. Effect modification was assessed through stratified analysis to identify variables whose effect, when combined with the level of nonoxynol-9 use, was greater than would be expected if the effects were independent and simply additive. Potential confounders of the exposure-disease relationship identified through stratified analysis were included in initial proportional hazards regression models and then dropped, based on the extent to which their exclusion changed the estimated rate ratios.[17] The rate ratios for the exposure variables were estimated by SAS proportional hazards regression software.[18]

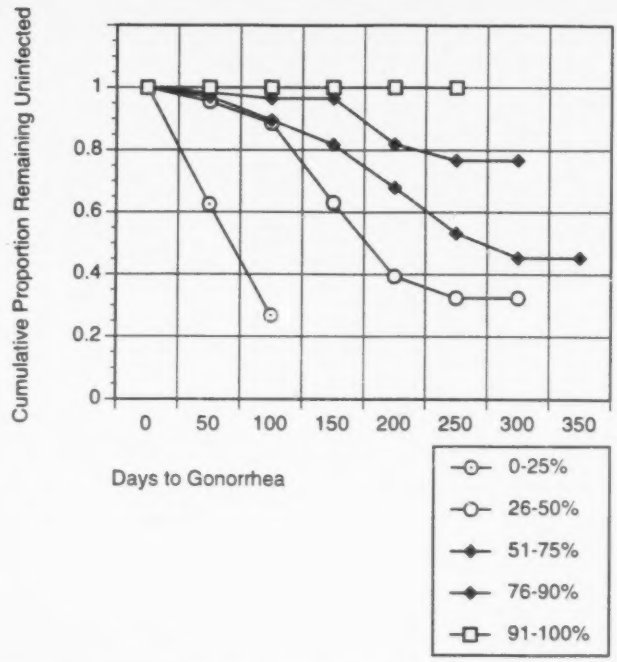
Results

Because the primary research question for the cohort study was the protective effect of nonoxynol-9 against HIV, 30 women were excluded from the cohort owing to positive ELISA tests and confirmatory Western blot, either at admission or during the first 3 months. Western blot results were considered positive if they showed at least two envelope bands (World Health Organization criteria[19]). Of the remaining 273 women, seven (2.6%) with positive cultures for endocervical gonorrhea at admission were excluded from the analysis, as was one woman with missing gonorrhea test data at baseline but confirmed evidence of gonorrhea at the time of the first follow-up. Three women did not report any sexual partners during the brief period they were in the study before dropping out; they were also excluded. Included in the analysis, then, were 241 women with negative gonorrhea culture tests at admission and 21 additional women whose gonorrhea culture results were missing at admission but who had a negative culture at the next follow-up. This latter group was treated as if their date of admission was the first follow-up visit.

At admission, approximately a third of the women were less than age 25; 37% were 25 to 29; and the remainder were 30 to 45. About 45% had completed schooling in addition to primary school. Fewer than 10% had ever married. The women were sexually active; over 40% had more than 15 sexual partners per month. Fewer than 15% reported using a contraceptive method at admission.

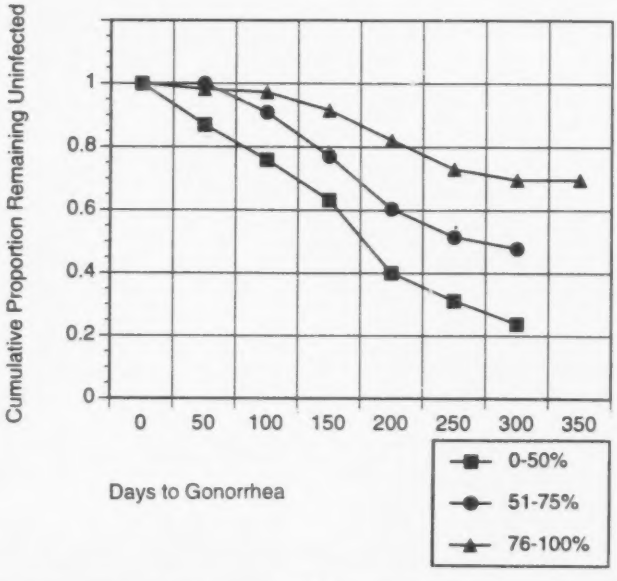
During follow-up, compliance in using condoms and nonoxynol-9 was quite high (Table 1): the mean percentage of coital acts protected by using both methods together was 40%; by using condoms alone, 22%; by using nonoxynol-9 alone, 25%; and by using neither, only 12%. Eight percent reported using both condoms and nonoxynol-9 for more than 75% of all acts; 22% used either condoms, nonoxynol-9, or a combination for all acts; and fewer than 2% used no protection at all for more than 50% of all acts.

Ninety-five women contracted gonorrhea during the almost 1500 person-months of observation for an estimated incidence of 6.2 gonorrhea infections per 100 person-months of observation. Of the remaining 167 women, 94 were censored because of study discontinuation, 23 were censored because of incomplete data, and 50 were censored at the end of the study period. Initial life tables demonstrated the protective effect against gonorrhea of condoms (Figure 1; $P < .0001$ log-rank test) and of nonoxynol-9 during acts not protected by condoms (Figure 2; $P < .0001$ log-rank test). The incidence of gonorrhea by nonoxynol-9 level among high, medium, and low condom users (Table 2) and rate ratios show that the rate of infection fell as condom use increased and, within levels of condom



Note. Ranges represent levels of the variable (number of coital acts with condom use divided by the total number of coital acts reported).

FIGURE 1—Gonorrhea life table rates, by condom use.



Note. Ranges represent levels of the variable (number of coital acts with nonoxynol-9 use divided by the number of coital acts without condom use).

FIGURE 2—Gonorrhea life table rates, by nonoxynol-9 use during coital acts not protected by condoms.

use, as the level of nonoxynol-9 use increased.

Among the variables considered as potential effect modifiers or confounders during stratified analysis or initial proportional hazards regression (i.e., age, educational attainment, oral contraceptive use, previous treatment for sexually transmitted disease, and presence of a dependent child), none clearly modified the dose–response protective effect of nonoxynol-9 described in Table 2. At levels of use higher than 50%, the protective effect of nonoxynol-9 was lower for a subgroup of young women (younger than age 30) who had completed secondary school than it was for other women; however, this was not evident at lower levels of nonoxynol-9 use. Because there was not a consistent difference in effect at all levels of nonoxynol-9 use and because the joint effect of being a young educated woman and of nonoxynol-9 level was less than multiplicative, a variable indicating membership in this subgroup was not included in the model.

The rate ratios and 95% confidence intervals (Table 3) were estimated by a proportional hazards model that included three levels of condom use (with less than 50% use being the reference group) and three levels of nonoxynol-9 use (with less than 25% use being the reference group). After controlling for condom use, using nonoxynol-9 during acts not protected by condoms provided increasing protection with increasing level of suppository use. Women using nonoxynol-9 for more than 75% of acts not protected by condoms had a rate of infection that was 14% that of women who used the compound for fewer than 25% of such acts. Even using nonoxynol-9 for only 26% to 50% of acts not protected by condoms reduced the rate of infection by half.

Acts protected by condoms in the life table and proportional hazards analyses reported above included acts protected by both a condom and a nonoxynol-9 suppository as well as acts protected only by a condom. The strong protective effect of nonoxynol-9 that was demonstrated in these data, however, raised the issue of the relative protective effects of nonoxynol-9 suppositories, condoms, and the concurrent use of both methods. Results from a proportional hazards model that included four continuous variables—the proportion of acts protected by condoms alone; the proportion protected by nonoxynol-9 alone; the proportion protected by both methods; and the number of sexual acts during the period—showed no

significant difference between the protective effect of using condoms alone, of using nonoxynol-9 alone, or of using both together. For every 1% increase in the use of condoms alone, nonoxynol-9 alone, or both together, the risk of gonorrhea was reduced. The estimated rate ratios (and 95% confidence intervals) were 0.956 (0.942, 0.971) for both methods together; 0.963 (0.946, 0.981) for condoms alone; and 0.960 (0.943, 0.976) for nonoxynol-9 suppositories alone. The protective effect of each method was also apparent in other exploratory proportional hazard models that used the number rather than the proportion of acts protected by each method.

The single best predictor in these data of the rate of gonorrhea was the proportion of acts protected by neither nonoxynol-9 nor condoms. Rate ratios estimated by a proportional hazards model that included only the proportion of acts not protected by either method (Table 4) show that as the proportion of unprotected acts increased, the rate of infection increased sharply.

TABLE 2—Incidence Rates of Gonorrhea per 100 Person-Months and Rate Ratios (RRs), by Levels of Condom Use and Nonoxynol-9 Use

Level of Nonoxynol-9 Use	Level of Condom Use: ≤50%		51%–75%		>75%	
	Incidence	RR	Incidence	RR	Incidence	RR
<25%	39.6	1.00	12.9	1.00	10.4	1.00
26%–50%	20.3	0.51	12.1	0.94	4.5	0.43
51%–75%	10.6	0.27	6.6	0.51	1.8	0.17
>75%	7.4	0.19	3.5	0.27	1.2	0.12

Note. Level of condom use refers to the percentage of total coital acts in which a condom or both a condom and nonoxynol-9 was used. Level of nonoxynol-9 use refers to the percentage of the remaining acts that were not protected by a condom or by both a condom and nonoxynol-9 and in which only a nonoxynol-9 suppository was used.

TABLE 3—Rate Ratios for Gonorrhea Infection from Proportional Hazards Model of Nonoxynol-9 Use, with Condom Use Controlled

	Rate Ratio	95% Confidence Interval
Level of nonoxynol-9 use among acts not protected by condoms		
>75%	0.14	0.07, 0.29
51%–75%	0.23	0.12, 0.46
26%–50%	0.48	0.23, 0.97
0%–25%[a]	1.00	. . .
Level of condom use		
>90%	0.05	0.01, 0.39
76%–90%	0.20	0.10, 0.39
51%–75%	0.44	0.29, 0.69
0%–50%[a]	1.00	. . .

[a]Reference group.

Discussion

This study confirms findings from other studies that have shown that condoms offer significant protection against gonorrhea. Given the risk of other sexually transmitted diseases, including HIV infection, women who have multiple sexual partners, such as the sex workers enrolled in this study, should consistently use condoms. However, in cases where condoms are not used owing to refusal by the male partner or to other reasons, this study demonstrates the substantial protection against gonorrhea afforded by nonoxynol-9 and provides evidence that use of the suppository alone is a reasonable strategy, especially when the alternative is a lack of protection by either method.

The level of protection provided by nonoxynol-9 in this study is similar to that reported in other studies. In fact, the rate ratios comparing gonorrheal infection among high and moderate nonoxynol-9 users with infection among less frequent users are quite similar. Niruthisard et al.[15] compared the incidence rates for chlamydial or gonococcal cervical infections among three categories of nonoxynol-9 users based on the percentage of coital acts in which nonoxynol-9 was used (76% to 100%, 51% to 75%, and 0% to 50%). They reported rate ratios of 0.4 and 0.7 for the high- and moderate-use groups, respectively, compared with the lowest-use group. Using similar categories (76% to 99% use, 51% to 75% use, and 0% to 50% use), Louv et al.[13] reported rate ratios of 0.3 and 0.5 for the high- and moderate-use groups, respectively. When the categories of Niruthisard et al. were applied to the data from Cameroon, the estimated rate ratios were similar: 0.4 for high users and 0.6 for moderate users compared with the lowest-use group.

Previous studies have not taken into account the extent to which women used condoms and nonoxynol-9 together or separately. This may have limited the ability of these studies to detect the dose–response protective effect of nonoxynol-9 evident in these data. This is because women who use nonoxynol-9 as a backup method when condoms are not used may have fewer completely unprotected acts (but lower absolute nonoxynol-9 use) than women who use condoms and nonoxynol-9 together most, but not all, of the time. In this study, the best indicator of risk of gonorrhea was the proportion of acts left unprotected by either method, not the absolute level of condom or nonoxynol-9 use.

TABLE 4—Rate Ratios for Gonorrhea Infection from Proportional Hazards Model of Proportion of Coital Acts Not Protected by Condoms or Nonoxynol-9

Proportion of Acts Unprotected	Rate Ratio	95% Confidence Interval
0–9%[a]	1.0	. . .
10–19%	2.6	1.5, 4.5
20–29%	4.5	2.5, 8.1
30–39%	9.0	4.6, 17.6
≥40%	31.0	13.3, 72.1

[a]Reference group.

There are several weaknesses of this study. The lack of randomization to condom or nonoxynol-9 use means that there may be unknown factors associated with risk of gonorrhea that differ substantially in their distribution between users of condoms and of nonoxynol-9. Perhaps young educated women who were high users of nonoxynol-9 were more likely to have partners with a higher prevalence of gonorrhea; perhaps women who used both methods at every intercourse, and were therefore more compliant, were more eager to reduce their risk of infection in other ways and avoided other unknown activities that increased the risk of infection.

In addition, the level of protection reported here for both condoms and nonoxynol-9 is likely to be underestimated in that the level of condom and nonoxynol-9 use among the reference groups was still significantly higher than zero. To the extent that condom and nonoxynol-9 use were overreported, which would not be surprising in spite of the admonitions given to record use of both methods accurately, the rate ratios would also be underestimated. The level of protection reported here is likely to be overestimated for the entire population of women from whom the cohort was obtained in that the women who volunteered for this study probably have higher levels of compliance with condom and nonoxynol-9 use than women who were not able to complete monthly coital logs and attend the clinic regularly for monthly checkups.

Results from the proportional hazard models exploring the relative protection afforded by nonoxynol-9 alone, condoms alone, and both methods together should be interpreted with caution. The estimated rate ratio for a 20% increase in the use of nonoxynol-9, for example, is 0.44, indicating that the risk of gonorrhea would be reduced by more than half. Even though there was no clear evidence of the risk of gonorrhea being nonlinear in the logit for the continuous variables, we are not willing to assume that an increase from 30% to 50% would have the same protective effect as an increase from 80% to 100%.

In conclusion, although the protective effect of condoms against sexually transmitted disease is greater than that afforded by nonoxynol-9, the protective benefit of nonoxynol-9 against gonorrhea is substantial. Using nonoxynol-9 when condoms are not used is a far preferable strategy in gonorrhea prevention than using no method at all. As Cates and Hinman argue in the case of HIV prevention,[20] controlling the gonorrhea epidemic will require different, mutually reinforcing approaches and the use of partially effective methods, which, if more acceptable to the target population, may ultimately have a greater effect on disease incidence than may more effective but less well accepted methods. □

Acknowledgments

Partial support for this work was provided by Family Health International (FHI) with funds from the US Agency for International Development. FHI is an international nonprofit organization that conducts research and provides technical assistance in health, family planning, sexually transmitted diseases, and AIDS.

The authors acknowledge the contributions of Dr Gina Dallabetta, Dr Judith Fortney, Mr Tom Grey, and Ms Barbara Taylor of Family Health International; Dr Irva Hertz-Picciotto of the Department of Epidemiology, University of North Carolina at Chapel Hill; and the three anonymous reviewers of this article.

References

1. Roper WL, Peterson HB, Curran JW. Commentary: condom and HIV/STD prevention—clarifying the message. *Am J Public Health.* 1993;83:501–503.
2. Cates W Jr, Stone KM. Family planning, sexually transmitted disease and contraceptive choice: a literature update. *Fam Plann Perspect.* 1992;24:75–84.
3. Cates W Jr, Stewart FH, Trussell J. Commentary: the quest for women's prophylactic methods—hopes vs science. *Am J Public Health.* 1992;82:1479–1482.
4. Centers for Disease Control and Prevention. Update: barrier protection against HIV infection and other sexually transmitted disease. *MMWR.* 1993;42:589–591, 597.
5. Weller SC. A meta-analysis of condom effectiveness in reducing sexually transmitted HIV. *Soc Sci Med.* 1993;36:1635–1644.
6. Stein Z. HIV prevention: the need for methods women can use. *Am J Public Health.* 1990;80:460–462.
7. Rosenberg MJ, Gollub EL. Commentary: methods women can use that may prevent sexually transmitted disease, including HIV. *Am J Public Health.* 1992;82:1473–1478.
8. Judson FN, Ehret JM, Bodin GM, Levin MJ, Rietmeijer CAM. *In vitro* evaluations of condoms with and without nonoxynol-9 as physical and chemical barriers against *chlamydia trachomatis,* herpes simplex virus type 2, and human immunodeficiency virus. *Sex Transm Dis.* 1989;16:51–56.
9. Rietmeijer CAM, Krebs JW, Feorino PM, Judson FN. Condoms as physical and chemical barriers against human immunodeficiency virus. *JAMA.* 1988;259:1851–1853.
10. Feldblum PJ, Fortney JA. Condoms, spermicides, and the transmission of human immunodeficiency virus: a review of the literature. *Am J Public Health.* 1988;78:52–54.
11. Jick H, Hannan MT, Stergachis A, Heidrich F, Perera DR, Rothman KJ. Vaginal spermicides and gonorrhea. *JAMA.* 1982;248:1619–1621.
12. Austin H, Louv WC, Alexander WJ. A case-control study of spermicides and gonorrhea. *JAMA.* 1984;251:2822–2824.
13. Louv WC, Austin H, Alexander WJ, Stagno S, Cheeks J. A clinical trial of nonoxynol-9 for preventing gonococcal and chlamydial infections. *J Infect Dis.* 1988;158:518–523.
14. Rosenberg MJ, Rojanapithayakorn W, Feldblum PJ, Higgins JE. Effect of the contraceptive sponge on chlamydial infection, gonorrhea, and candidiasis: a comparative clinical trial. *JAMA.* 1987;257:2308–2312.
15. Niruthisard S, Roddy RE, Chutivongse S. Use of nonoxynol-9 and reduction in rate of gonococcal and chlamydial cervical infections. *Lancet.* 1992;339:1371–1375.
16. Zekeng L, Feldblum PJ, Oliver RM, Kaptue L. Barrier contraceptive use and HIV infection among high-risk women in Cameroon. *AIDS.* 1993;7:725–731.
17. Mickey RM, Greenland SA. A study of the impact of confounder-selection criteria on effect estimation. *Am J Epidemiol.* 1989;129:125–137.
18. *SAS/STAT Software: The PHREG Procedure, Version 6.* Cary, NC: SAS Institute Inc; 1991. SAS Technical Report P-217.
19. World Health Organization. Acquired immunodeficiency syndrome: proposed WHO criteria for interpreting results from Western blot assays for HIV-1, HIV-2 and HTLV-1/HTLV-II. *Weekly Epidemiol Rec.* 1990;65:281–283.
20. Cates W, Hinman A. AIDS and absolutism—the demand for perfection in prevention. *N Engl J Med.* 1992;327:492–494.

Are Bisexually Identified Men in San Francisco a Common Vector for Spreading HIV Infection to Women?

Maria L. Ekstrand, PhD, Thomas J. Coates, PhD, Joseph R. Guydish, PhD, MPH, Walter W. Hauck, PhD, Linda Collette, MS, and Stephen B. Hulley, MD, MPH

ABSTRACT

Objectives. This article examines sexual risk taking among self-identified bisexual men in San Francisco and whether risk reduction has occurred, with respect to both homosexual and heterosexual behaviors, among human immunodeficiency virus (HIV) antibody–positive and HIV antibody–negative men. It also examines psychosocial correlates of unprotected anal and vaginal intercourse.

Methods. The participants were members of a population-based longitudinal cohort of 1034 single men aged 25 through 54 years recruited from the 19 census tracts in San Francisco that had the greatest prevalence of the acquired immunodeficiency syndrome in 1984. Of the total sample, 140 subjects initially identified themselves as bisexual; 85% of these men remained in the study.

Results. The participants reported dramatic reductions in sexual risk taking. Prevalences of unprotected anal sex with men were similar among HIV-positive bisexual men (89% in 1984–1985 and 18% in 1988–1989) and those who were HIV negative (65% and 20%). The prevalence of unprotected vaginal sex was much lower for HIV-positive men (16% in 1984–1985 and 2% in 1988–1989) than for HIV-negative men (35% and 20%). Unprotected intercourse was associated primarily with situational and interpersonal factors.

Conclusions. Striking reductions in risk behaviors were reported. This subgroup of single, bisexually identified men appears unlikely to be a common vector for spreading HIV infection to women. (*Am J Public Health.* 1994;84:915–919)

Introduction

This study was designed to examine the prevalence of and change in sexual behaviors among San Francisco men who identified themselves as bisexual. The US Department of Health and Human Services reported that as of October 1991, 10% (n = 625) of all heterosexually transmitted acquired immunodeficiency syndrome (AIDS) cases among women were reported to be due to sexual contact with a bisexual man.[1] It is important to learn more about both the prevalence and the patterns of sexual behaviors among bisexual men, as well as psychosocial correlates of unprotected sex, as a guide to developing appropriate AIDS prevention campaigns for this population.

A recent study that used national surveillance data on US men with AIDS found that at least 14% were behaviorally bisexual.[2] The results further showed that nearly one quarter of bisexual men with AIDS who died were married at the time. It is unclear to what extent female sex partners are aware of the behaviors of bisexual men, but data from the California Partners' Study[3] suggest that this awareness may vary depending on ethnicity. Among 52 female partners of human immunodeficiency virus (HIV) antibody-positive bisexual men, 80% of the White, 20% of the African-American, and 22% of the Latina women were aware that their male sex partners were bisexual. Although these studies suggest that women may indeed be at risk for HIV transmission from bisexual men, they provide no data on the sexual risk-taking behaviors of these men.

In a study of homosexually active British men,[4] 61% of the participants reported having engaged in some heterosexual behaviors during their lifetime. Although only 12% of the sample reported having had a female partner during the past year, 92% of these men stated that these were "penetrative partners" (i.e., partners with whom they had practiced anal or vaginal sex). In addition, condom use was poor among these men, with only 28% reporting consistent condom use during vaginal intercourse. These findings suggest that there may be differences in sexual risk taking among bisexual men depending on the gender of one's sex partner. However, they provide no information about behavior change or correlates of risk taking.

The San Francisco Men's Health Study consists of single men who at baseline had engaged in sex with men, women, or both in the past 2 years. An early study[5] of the participants found that 5% (n = 50) of the men were HIV positive and bisexual. These men reported having had sex with an average of 2.9 women during the 2 years before baseline. This finding led the authors to conclude that "a substantial potential clearly exists for epidemic spread of [HIV] infection to women in San Francisco."[5(p901)]

In our previous studies of behavior change in this sample, we found that gay men made unprecedented and lasting sexual-behavior changes.[6] No analysis has been conducted of bisexual behavior since the 1984 baseline visit.[5] Previous research[6–9] with this and other cohorts

The authors are with the Center for AIDS Prevention Studies, University of California, San Francisco, Calif.

Requests for reprints should be sent to Maria L. Ekstrand, PhD, UCSF Center for AIDS Prevention Studies, S-600, 74 New Montgomery, San Francisco, CA 94105.

This paper was accepted September 30, 1993.

identified several important correlates of unprotected anal intercourse among gay men. We are not aware of any previous studies of correlates of sexual risk taking among bisexual men.

This article focuses on sexual behavior and its psychosocial correlates among self-identified bisexual men in the San Francisco Men's Health Study. This analysis had four goals: (1) to examine the relationship between behavioral bisexuality and having a bisexual identity; (2) to determine whether risk reduction has occurred with respect to both homosexual and heterosexual activities within this population; (3) to examine the relationship between HIV antibody status and unprotected sexual intercourse among bisexual men; and (4) to determine whether there is an association between psychosocial variables and unprotected sexual intercourse with men and/or women in this population.

Methods

Subjects and Procedures

The San Francisco Men's Health Study is a prospective study of single men recruited by stratified two-stage probability sampling from the 6-km^2 area in San Francisco with the highest cumulative incidence of AIDS in 1984. The study sample and methods were described in detail previously.[10,11] All single men between the ages of 25 and 54 years who lived in one of the selected housing units were contacted by the survey workers and invited to participate in the study. Fifty-nine percent (n = 1034) of these men entered the study. Analyses of selected characteristics of the study subjects and those of comparable men in the same census districts showed the study sample to be representative of all contacted subjects with regard to demographic characteristics.[11] We have no data on whether study participants were representative of all contacted men in terms of their sexual behaviors or of bisexual men who are married or live in other parts of San Francisco.

The initial wave of data collection took place in July 1984 at the outpatient clinic of the Children's Hospital in San Francisco. The interviews lasted approximately 1 hour and focused on sexual behaviors, physical symptoms, other relevant health-related behaviors, and psychosocial variables. After the interview, a physical examination was performed by specially trained physician assistants. Follow-up assessments have taken place every 6 months. Psychosocial questionnaires were administered in 1985, 1987, and 1989.

Participants were given a list of different terms describing sexual orientation (i.e., gay, homosexual, bisexual, straight, heterosexual, and "other"). All participants who described themselves as bisexual in the first year of data collection (n = 140) were included in these analyses. Demographic subject characteristics obtained in 1984 showed that 85% of the bisexual men were White, 7% were Black, and 7% were Latino. The age range at entry into the study was 25 to 54 years, with a mean of 35 years. The majority of the participants (64%) had bachelor's degrees, and 56% of these had completed some postgraduate work. These rates were comparable to those obtained for the entire San Francisco Men's Health Study sample.[6]

Measures

Sexual behavior. Self-report data on high-risk sexual behavior were obtained every 6 months, starting in July 1984 and ending in June 1989. In the present analyses, we examined the number of male and female sex partners, unprotected anal and vaginal sex, and oral sex with men and women. The semiannual data from the first and last two waves were combined into two 12-month segments, labeled "year 1" and "year 5." Thus, an individual who had practiced a behavior in either or both of two consecutive 6-month waves would be coded as having practiced the behavior in that year.

Because these were secondary data analyses, the selection of measures of potential risk-behavior correlates was not theoretically based. Rather, our decision to include these measures was primarily empirically based. Virtually all of these measures were previously shown to be associated with risk-taking behaviors in cohorts of gay and bisexual men.

Age. Age was entered as a continuous variable, ranging from 25 to 54 years in 1984.

HIV antibody status. Participants' blood samples were examined independently by two laboratories with an indirect immunofluorescent antibody assay. Antibody-positive results were confirmed by using immunoblot tests. Baseline data included blood tests obtained in 1984 and 1985 (frozen sera from 1984 were tested in 1985), whereas follow-up data included HIV antibody status through 1989. There were seven seroconverters in this sample during the study period. Men who chose to receive HIV antibody results received standard posttest counseling. Previous analyses have shown that knowledge of antibody status did not influence risk behaviors in this cohort.[12]

Alcohol consumption. Data on alcohol consumption were obtained in 1989 for the respondents' drinking behavior during the previous 6 months. This scale was constructed to show the average volume of alcohol consumption per week. The mean amount of alcohol consumption in this sample was seven drinks per week, with scores ranging from 0 to 53. This scale was entered as a continuous variable, ranging from 0 to 53.

HIV-related health beliefs. Three types of beliefs were studied: perceived self-efficacy (four items, $\alpha = .58$), HIV transmission-related beliefs (two items, $\alpha = .45$), and perceived susceptibility to HIV/AIDS (two items, $\alpha = .85$). Participants were asked to indicate the amount of agreement with each item on the first two scales by using a 4-point scale ranging from "strongly agree" to "strongly disagree." Please see the Appendix for a list of health belief items and wording of response options.

Stress. This scale was designed to measure perceived stress in the past year and was administered in 1989. Areas assessed included perceived stress relating to the health of oneself and others, work, relationships, and finances. Participants rated each of the items on a four-point scale ranging from "none" to "a great deal" of stress (eight items, $\alpha = .61$).

Procedure

Attrition bias analyses were conducted by examining baseline rates for those subjects who remained in the study vs those who dropped out between 1984 and 1989. Dichotomous data were analyzed by using chi-square tests, whereas continuous data were analyzed with *t* tests.

The proportion of participants who identified themselves as bisexual at each of the study waves was compared with the proportion of men who reported having sex with both men and women at these times to examine the relationship between behavioral bisexuality and bisexual identity.

Sexual behavior prevalence rates were examined separately for HIV-positive and HIV-negative bisexual men in years 1 and 5. Sexual behaviors with men and women are described separately.

The association between the potential correlates and unprotected sex was examined by using backward stepwise multiple logistic regression, while controlling for age. The backward procedure was used to control for possible intercorrelations among the variables. Previous research[6] showed that age is an important predictor of high-risk sex among gay men in this sample. Controlling for age thus ensured that any other significant correlates contributed above and beyond the effect of age. For all scales, we compared scores separated by a fourth of the scale range (e.g., an interval of 3 points on a 12-point scale).

Results

Attrition Bias

We compared the participants seen in both 1984 and 1989 (n = 119) with the participants who dropped out (n = 21). No statistically significant differences were found. Both groups reported comparable rates of sexual behaviors in 1984–1985, with 78% and 27% of panel members reporting unprotected sex with men and women, respectively. The corresponding rates were 76% and 19% for dropouts. The mean age for both groups was 35 years. Forty-two percent of panel members and 43% of dropouts reported HIV-related symptoms. The two groups consumed similar amounts of alcohol (the mean number of drinks per week was 9.2 and 8.4, respectively). Forty-nine percent of the panel members and 67% of the dropouts were HIV positive.

Bisexual Behaviors and Identity

In year 1 of this study, 140 men identified themselves as bisexual. Thirty-seven of these men were also behaviorally bisexual during that time. An additional two men reported having sex with both men and women in year 1 but did not identify themselves as bisexual. In year 5 of this study, 119 men remained in our original cohort of bisexually identified men. Eight of these men were also behaviorally bisexual during that time. An additional three men reported having sex with both men and women in year 5 but did not identify themselves as bisexual.

Unprotected Anal and Vaginal Intercourse

All subject groups made significant risk reductions during this time period. The main differences appeared when comparing the gender of sex partners separately for HIV-positive and HIV-negative bisexual men. In 1984–1985, 89% of the HIV-positive men reported practicing unprotected anal sex with men, whereas 16% reported unprotected vaginal sex. By 1988–1989, 18% continued to report unprotected anal intercourse with men, and only 2% reported any unprotected vaginal sex with women.

Among HIV-negative men, a somewhat smaller proportion (65%) started out practicing unprotected anal sex with men. By year 5, this proportion was reduced to 20%, a figure comparable to that for HIV-positive men. The risk reduction was somewhat less pronounced for sex with women. In 1984–1985, 35% of these men reported engaging in unprotected vaginal sex, compared with 20% in 1988–1989. This means that in 1988–1989, 80% of all HIV-negative bisexual men who reported having sex with women (25% of the sample) reported having unprotected vaginal sex. This is a marked difference from their risk pattern with other men in 1988–1989. At that time, only 26% of the HIV-negative men who had sex with men (76% of sample) reported that they had engaged in unprotected anal intercourse.

Other Sexual Behaviors

The proportions of oral sex with men remained fairly constant among both HIV-positive and HIV-negative bisexual men. The percentage of HIV-positive men reporting this behavior decreased from 91% in 1984–1985 to 82% in 1988–1989. Among HIV-negative men it decreased from 83% to 76% during this time period. Slightly greater changes were seen in the percentage of men who reported oral sex with women. Among HIV-positive men, this percentage decreased from 23% to 7% and for HIV-negative men it decreased from 29% to 22%.

The percentage of men who reported sex with multiple (i.e., two or more) partners during a given year also decreased during the study period. The percent of HIV-positive men who reported multiple male partners decreased from 86% to 58%, whereas the percentage who had multiple female partners decreased from 7% to 2%. The percent of HIV-negative bisexual men with multiple male partners decreased from 63% to 51%, whereas the corresponding percentage with multiple female partners decreased from 21% to 8%. Finally, the prevalence of celibacy among all bisexual men increased from 4% to 13% during these 5 years.

Table 1 shows the proportions of participants who engaged in two different sexual behaviors during the same year in 1984–1985 and 1988–1989. The most drastic decrease occurred among men who engaged in unprotected anal intercourse and had multiple partners in year 1 (70%) and in year 5 (12%). In 1988–1989, very few (7%) men reported sex with men and women in the same year, compared with 26% in 1984–1985. The proportion of bisexual men who practiced unprotected sex with both men and women in the same year decreased from 16% to 2% of the sample.

TABLE 1—Percentage of Bisexual Men Who Reported Engaging in Multiple Sexual Behaviors during the Previous 12 Months (n = 119)

Sexual Behavior	Year 1 (1984–1985) %	Year 1 (1984–1985) No.	Year 5 (1988–1989) %	Year 5 (1988–1989) No.
Gender of sex partner				
Sex with men only	65	77	70	83
Sex with women only	6	7	10	12
Sex with both men and women	26	31	7	8
Celibate	5	6	13	15
Unprotected vaginal sex and multiple female partners	13	15	3	4
Unprotected anal sex with men and multiple male partners	70	83	12	14
Unprotected anal sex with women and multiple female partners	4	5	1	1
Unprotected anal sex with men and unprotected vaginal sex with women	16	19	2	2

TABLE 2—Cross-Sectional Correlates of Unprotected Anal and/or Vaginal Intercourse with Men or Women in 1989

Correlates	Odds Ratio	95% Confidence Interval
Age (10-y difference)	0.5	0.2,1.3
Self-efficacy	8.8	3.1,25.3
Susceptibility	1.3	0.6,2.8
Transmission beliefs	0.5	0.2,1.3
Alcohol consumption	1.1	0.6,1.9
Stress	0.7	0.3,1.5
HIV antibody status (positive)	0.6	0.1,4.8

Note. The overall Hosmer-Lemeshow goodness-of-fit chi-square test for the final step in the backward stepwise logistic regression (which included only self-efficacy after adjusting for age) was 68.79 ($P = .485$). The P values for the chi-square test of the coefficients at this step were $<.0001$ for self-efficacy and .3034 for age. The odds ratios reported above are from the first step in the backwards procedure, which included all of the variables in the equation.

Cross-Sectional Correlates of Unprotected Sex

Table 2 shows the six variables entered into the multiple stepwise backward logistic regression analysis predicting unprotected intercourse in 1989. Due to small numbers, we decided to combine unprotected intercourse with men and women into one risk variable. High risk was thus defined as reporting any unprotected anal sex with men or women and/or reporting unprotected vaginal sex during the previous 6 months. Subjects were classified as low risk if they reported no unprotected sex with men or women during this time period. Perceived self-efficacy was the only significant correlate of unprotected sexual intercourse. High-risk subjects reported being significantly less confident of their ability to practice safe sex in a variety of situations than did low-risk subjects. Once perceived self-efficacy was entered into the stepwise model, it remained as a significant correlate. No other variables were statistically significant at any of the steps. However, although the remaining variables were not statistically significant, their wide confidence intervals suggest that it may be premature to exclude them as potential correlates of unprotected sexual intercourse.

Hypothetical Risk Situations

Participants were also asked to indicate retrospectively how likely they were to have anal sex without a condom with another man in nine hypothetical situations. The following situations were endorsed as either "very likely" or "somewhat likely" by the indicated percentage of study participants: sex while under the influence of alcohol or other drugs (37%), being in love with one's sex partner (35%), not having any condoms available (33%), having a partner who doesn't want condoms (30%), having sex with one's lover (29%), planning to withdraw before ejaculating (25%), perceiving condoms as making sex less enjoyable (23%), being under a great deal of stress (19%), and being sad or upset (18%).

Discussion

Single bisexual men in San Francisco have made dramatic changes in their sexual behavior. Some interesting differences emerged when we considered HIV antibody serostatus. Among HIV-positive bisexual men, most of the sexual risk taking appeared to occur with other men, with levels of unprotected anal sex being reduced from 89% in 1984–1985 to 18% in 1988–1989. Only 7% of the men in this group reported that they had engaged in any form of sexual activity with women in 1988–1989. Two percent reported having had unprotected vaginal intercourse, and no unprotected anal sex with women was reported in this group in 1988–1989. These changes are very encouraging. However, given that the HIV seroprevalence rates are approximately 50% among gay men in this age group in San Francisco, any unprotected sexual activity with a male partner of unknown HIV antibody status carries a high risk of HIV transmission.

HIV-negative men continued to be approximately three times more likely to engage in some form of sexual activity with men than with women in 1988–1989. Among those who did report sexual activity with women in year 5, unprotected vaginal sex appeared to be the norm, with 80% of these men stating that they had engaged in unprotected vaginal intercourse. In contrast, only 26% of those who had any kind of sex with men practiced unprotected anal sex. Possibly, bisexual men who are not infected with HIV perceive sex with women as less risky, thus representing an alternative to condom use during male-to-male sex as a method of HIV risk reduction.

It is unclear whether HIV antibody testing influenced behaviors in this study. Based on the data from the first wave of data collection (1984), it appears that men who later tested HIV negative had higher rates of sex with women (37%) and slightly lower rates of sex with men (83%) than did men who later tested HIV positive (23% and 93%, respectively). It is thus possible that their prior behavior may have contributed to the risk of infection. The decreased rates of sex with women (from 23% to 7%) and unprotected vaginal sex (from 16% to 2%) among HIV-positive men suggest that at least some of the HIV-positive men may have modified their sexual behaviors with women after their HIV antibody testing. It is important to note that these hypotheses are tentative and need to be rigorously tested in future studies before any firm conclusions can be drawn regarding the effect of HIV antibody testing in this population.

There was a striking reduction in the percentage of men who had unprotected anal sex and multiple male partners during the 5 years of the study (from 71% to 12%). Also, significantly fewer men reported having both male and female sex partners in 1988–1989 than in 1984–1985 (26% vs 7%). It is possible that much of this change is due to an increased tendency to have longer relationships and fewer sex partners. We were unable to examine this possibility directly because we have very limited data on relationships with men and no such data on female sexual partners. Because this issue has important implications for AIDS prevention campaigns, it needs to be examined in future research.

The analyses of psychosocial variables suggest that much risk taking is situationally determined. The self-efficacy scale included items assessing the amount of confidence the participants had in their ability to practice safe sex in different situations, such as "in the heat of passion" and "with an uncooperative partner." It should be noted that because these are cross-sectional analyses, it is possible that the participants were simply labeling their behavior during the past year when responding to these items. However, low self-efficacy has previously been found to predict sexual risk taking longitudinally in another San Francisco cohort of gay and bisexual men[7] and has also been associ-

ated with a variety of other health-related behaviors.[13,14]

The retrospective reports provided additional information regarding potential high-risk situations. The most frequently endorsed situations involved having sex under the influence of alcohol and other drugs, condom availability, and relationship variables such as one's feelings for and requests by a sexual partner. These results are very similar to results from other cohorts of gay and bisexual men.[15] The consistency of these findings suggests that some bisexual men may not perceive themselves as possessing the skills necessary for safe sex negotiation in high-risk situations. Finally, it is interesting that although alcohol consumption itself was not associated with unprotected sexual intercourse in the regression analysis, the combination of substance use (including use of alcohol) and sexual activity was perceived by the participants as an obstacle to safe sex. It thus appears that the role of alcohol and other drugs in connection with sexual activity needs to be addressed in AIDS prevention programs, even though we may not yet completely understand the nature of this association.

It is important to note that this study has some limitations. These results may not generalize to San Francisco men who have sex with both men and women but who do not identify themselves as bisexual. San Francisco Men's Health Study participants appear likely to be more open about their sexual orientation than other samples of perhaps more closeted bisexual men.[16] In spite of these differences, we find our results to be very encouraging and an example of how risk reduction may be facilitated in a culture in which most behaviorally bisexual men also identify themselves as bisexual.

Another limitation of this study was its small sample size, which prevented us from examining correlates of risk taking separately for the gender of the sex partner. It is thus possible that different factors influence risk taking during male-to-male and male-to-female sex.

In summary, the study population has drastically reduced sexual risk behaviors, especially during male-to-male sexual activities. Although unprotected vaginal intercourse is still the norm among those men who have sex with women, this behavior was reported almost exclusively by HIV-negative men. Thus we conclude that single, bisexually identified men in San Francisco do not appear to be a common vector for spreading HIV disease to women. □

Acknowledgments

Preparation of this manuscript was supported in part by National Institutes of Health contract N01-A1-82515, by the National Institute of Mental Health/National Institute of Drug Abuse Center for AIDS Prevention Studies grant MH 42459, and by a scholarship grant to Dr. Ekstrand from the American Foundation for AIDS Research.

The authors thank Drs Jay Paul and Ron Stall of the Center for AIDS Prevention Studies and Mr Peter Weatherburn of the University of Essex, UK, for their critiques of an earlier version of this manuscript.

References

1. Centers for Disease Control. Table 3. AIDS cases by age group, exposure category, sex, and year of report. *HIV/AIDS Surveill.* September 1991: 1–18.
2. Chu SY, Peterman TA, Doll LS, Buehler JW, Curran JW. AIDS in bisexual men in the United States: epidemiology and transmission to women. *Am J Public Health.* 1992;82:220–224.
3. Padian NS. Female partners of bisexual men. Presented at the CDC workshop on bisexuality and AIDS. October 1989; Atlanta, Ga.
4. Weatherburn P, Davis PM, Hunt AJ, Coxon APM, McManus TJ. Heterosexual behavior in a large cohort of homosexually active men in England and Wales. *AIDS Care.* 1990;2:319–324.
5. Winkelstein W, Wiley J, Padian N. Potential for transmission of AIDS-associated retrovirus from bisexual men in San Francisco to their female sexual contacts. *JAMA.* 1986;255:901.
6. Ekstrand ML, Coates TJ. Maintenance of safer sexual behaviors and predictors of risky sex: the San Francisco Men's Health Study. *Am J Public Health.* 1990;80:973–977.
7. McKusick L, Coates T, Morin S, Pollack L, Hoff C. Longitudinal predictors of reductions in high risk sexual behaviors among gay men in San Francisco: the AIDS Behavioral Research Project. *Am J Public Health.* 1990;80:978–983.
8. Stall R, McKusick L, Wiley J, et al. Alcohol and drug use during sexual activity and compliance with safe sex guidelines for AIDS: the AIDS Behavioral Research Project. *Health Educ Q.* 1986;13:359–371.
9. Hays R, Kegeles S, Coates T. High HIV risk-taking among young gay men. *AIDS.* 1990;4:901–907.
10. Winkelstein W, Lyman D, Padian N, et al. Sexual practices and the risk of infection with the human immunodeficiency virus. *JAMA.* 1987;257:321–325.
11. Piazza T. Sampling methods and wave 1 field results of the San Francisco Men's Health Study. A report to the National Institutes of Allergy and Infectious Disease. April 1986.
12. Coates TJ, Ekstrand ML, Kegeles SM, Stall RD. HIV antibody status and behavior change in two cohorts of gay men in San Francisco. Paper presented at the Fifth International Conference on AIDS; June 1989; Montreal, Canada.
13. Bandura A. Perceived self-efficacy in the exercise of control over AIDS infection. In: Mays V, Albee G, Jones J, Schneider S, eds. Psychological Approaches to the Prevention of AIDS. Beverly Hills, Calif: Sage Publications; 1989.
14. Bandura A. A social cognitive approach to the exercise of control over AIDS infections. In: DiClemente R, ed. Adolescents and AIDS: A Generation in Jeopardy. Beverly Hills, Calif: Sage Publications; 1991.
15. Stall R, Ekstrand M, Pollack A, McKusick L, Coates T. Relapse from safer sex: the next challenge for AIDS prevention efforts. *J Acquir Immune Defic Syndr.* 1990;3:1181–1187.
16. Lever J, Kanouse DE, Rogers WH, Carson S, Hertz R. Behavior patterns and sexual identity of bisexual males. *J Sex Research.* 1992;29:141–167.

APPENDIX—Health Belief Measures

1. Self-efficacy (agreement expressed on a 4-point scale)
   A. I am usually able to convince my sexual partner to have safe sex.
   B. If my partner refuses to use a condom, I still will have intercourse with him or her.
   C. I believe there is a lot I can do to prevent HIV transmission.
   D. In the heat of passion, I sometimes engage in sexual activities that are unsafe.
2. Susceptibility (degree of certainty expressed on 4-point scales, plus 1 N/A item [i.e., "I'm already infected" or "I've already been diagnosed"])
   A. How likely do you think it is that you will become infected with HIV in the future?
   B. How likely do you think it is that you will eventually get AIDS?
3. Transmission beliefs (agreement expressed on a 4-point scale)
   A. If both partners know that they have the antibodies (i.e., they both have positive test results), they can have any kind of sex without increasing their risk of getting AIDS.
   B. If both partners know that they do not have antibodies to the virus (i.e., they both have negative test results), they don't have to worry about practicing safe sex.

Correlates of Needle Sharing among Injection Drug Users

Wallace Mandell, PhD, David Vlahov, PhD, Carl Latkin, PhD, Maria Oziemkowska, MS, and Sylvia Cohn, MPH

ABSTRACT

Objectives. The sharing of contaminated injection equipment is the primary mode of human immunodeficiency virus (HIV) transmission for injection drug users. This study examined demographic factors, life events, and drug use practices that are potential risk factors for sharing injection equipment.

Methods. Between February 1988 and March 1989, 2921 active injection drug users were interviewed and questioned about their backgrounds, life-styles, and patterns of injection drug use.

Results. Of 2524 participants who reported injecting drugs within the 6 months prior to study enrollment, 70.4% reported recent needle sharing. A multivariate analysis found needle sharing to be more frequent among those with a history of arrest and lower socioeconomic status, even after accounting for other demographic and drug use variables. In addition, recent needle sharing was higher in male homosexual or bisexual men than in their heterosexual counterparts.

Conclusions. These data suggest that injection drug users have an economic motive to share needles and that the availability of free and legal needles may reduce levels of needle sharing. (*Am J Public Health.* 1994;84:920–923)

Introduction

Injection drug users are at increased risk for human immunodeficiency virus (HIV) infection. The primary mechanism for transmission of HIV infection among injection drug users is sharing contaminated needles and syringes.[1–4] Education about risk reduction and distribution of bleach for disinfecting needles have been important strategies against HIV infection in this population.[5,6] However, in one study, 98% of active injection drug users knew that HIV was transmitted through sharing needles, yet 70% reported sharing needles prior to the knowledge interview[7]; these data suggest that education is necessary but insufficient to effect behavior change. More recently, a study of the effectiveness of bleach on HIV seroconversion noted a modest protective effect for needle disinfection as it was being practiced in the field.[8] These results and others[9,10] underscore the need to study factors that contribute to needle sharing practices in order to identify characteristics that might lead to rational preventive interventions.

Information about risk factors for needle sharing is limited, and most of the information comes from participants in drug abuse treatment programs.[9,10] However, only a small proportion of injection drug users are in treatment at any given time, and characteristics of injection drug users in treatment differ from those of users who are not in treatment.[11] Corresponding information about factors associated with needle sharing is needed for the latter group. In 1988 and 1989, a large sample of injection drug users not in treatment was recruited through extensive community outreach techniques for a study of the natural history of HIV infection; we report here the correlates of needle sharing at time of entry into the study.

Methods

Study Population

The rationale, organization, and methods for the AIDS Link to Intravenous Experiences (ALIVE) Study have been described in detail elsewhere.[12] Briefly, between February 1988 and March 1989, injection drug users were recruited through extensive community outreach techniques. Brochures about the study were distributed at local drug abuse treatment programs, sexually transmitted disease clinics, emergency rooms, state parole and probation offices, homeless shelters, and subsidized housing projects. In addition, outreach workers from a local community health education organization distributed brochures, and study participants informed eligible contacts through word of mouth. Eligibility criteria were age older than 17 years, history of drug injection any time during the previous 10 years (to correspond to the interval when HIV was surmised to be present in the community), and no diagnosis of acquired immunodeficiency syndrome (AIDS) at baseline (as this was a longitudinal outcome).

Wallace Mandell, Carl Latkin, and Maria Oziemkowska are with the Department of Mental Hygiene, and David Vlahov and Sylvia Cohn are with the Department of Epidemiology, the Johns Hopkins School of Hygiene and Public Health, Baltimore, Md.

Requests for reprints should be sent to Wallace Mandell, PhD, Department of Mental Hygiene, The Johns Hopkins University, School of Hygiene and Public Health, 624 N Broadway, Baltimore, MD 21205.

This paper was accepted September 20, 1993.

Data Collection

After pretest counseling, informed consent, and venipuncture, participants underwent face-to-face interviews with a trained interviewer in a private room within the study clinic. The interviewer gathered information on medical history, drug use, and sex practices, as described elsewhere.[13] For this analysis we used sociodemographic information for the 10 years prior to interview, sexual orientation, and drug use practices during the 6 months prior to last use. Drug use practices included type and frequency of drug used, duration of injection drug use, the practice of "splitting" drugs (drawing up drugs in one syringe and squirting half into a second syringe, which is a practice used when two or more people with their own syringes buy drugs together). Needle sharing was defined by the question "With how many different people did you share a needle in the past 6 months?"; a response of zero defined a nonsharer and a response of one or more defined a needle sharer.

Following the interview, participants received risk reduction counseling, an appointment to return in 2 weeks for HIV antibody test results, and $10 compensation for time spent in the study. The study procedures were reviewed and approved by the Institutional Review Board of The Johns Hopkins School of Hygiene and Public Health.

Statistical Analyses

The binary variable, needle sharing, was cross-tabulated with other study variables; odds ratios were generated with 95% confidence intervals to guide interpretation. To simultaneously adjust for potentially confounding variables, we used logistic regression procedures.[14]

Results

Of the 3375 individuals who registered for the study, 378 (11.2%) did not qualify because they were younger than 18 years of age or did not report injection drug use in the previous 10 years. An additional 76 (2.5%) were excluded because of duplicate registration or disqualifying information obtained by further investigation. Of the 2921 eligible and consenting participants, 2597 (88.9%) reported using injection drugs within the 6 months prior to initial enrollment in the study; 2524 had complete information available. Characteristics of the 2524 with complete information and the 73 with at least one variable missing were statistically similar (data not shown).

TABLE 1—Demographic Factors Associated with Needle Sharing during 6 Months prior to Last Injection, Baltimore, Md, February 1988 through January 1990 (n = 2524)

	No. Participants	% Who Shared	Univariate OR (95% CI)	Adjusted OR (95% CI)[a]
Age				
18–44 y	2013	71.2	1.19 (0.96,1.46)	1.41 (1.12,1.77)
45+ y	511	67.1		
Race				
Non-Black	244	73.0	1.14 (0.85,1.54)	1.11 (0.81,1.52)
Black	2280	70.2		
Sexual orientation				
Female	473	69.3	0.99 (0.80,1.23)	1.05 (0.93,1.19)
Male bisexual	196	81.6	1.95 (1.33,2.83)	1.72 (1.17,2.53)
Male heterosexual	1855	69.5		
Education				
9 y	203	79.3	1.67 (1.17,2.37)	1.56 (1.08,2.25)
9+ y	2321	69.7		
Legal income (last year)				
$5000	1873	69.1	0.92 (0.76,1.11)	1.07 (0.87,1.33)
≤$4999	651	70.9		
Currently unemployed				
Yes	1983	71.6	1.28 (1.04,1.56)	1.08 (0.86,1.35)
No	541	66.4		

Note. OR = odds ratio; CI = confidence interval.
[a]From logistic regression model presented in text, which controlled for demographic, life event, and drug involvement variables.

The risk factor analyses presented here are based on the sample of 2524 individuals who were active injection drug users in that they reported at least one intravenous injection in the 6 months prior to their initial visit to the study and who had no missing data. This group can be characterized as young (80% were between the ages of 25 and 44 years), male (81%), Black (90%), and currently unemployed (79%), with a history of imprisonment (66%) and homelessness in the past 10 years (49%), and often with no experience of treatment for drug abuse (48%) (Tables 1 and 2).

The subjects examined were active injection drug users: 40% injected at least daily and an additional 35% injected at least once a week; only 8% injected less often than once a month. Nearly one third attended shooting galleries (i.e., clandestine locations where needles and syringes are rented repeatedly). The predominant drug of choice for injection was cocaine (92%), although 77% used multiple drugs. Of the men, 9.6% engaged in homosexual or bisexual behavior and two thirds of these individuals, 6.4% of the men, had engaged in receptive anal intercourse within the last 10 years.

In univariate analysis (Tables 1 and 2), needle sharing was associated with almost all demographic (age and education), life-style (employment, public assistance, arrest), and drug involvement variables (drug of choice, frequency of injection, duration of use, experience of withdrawal symptoms, and participation in drug treatment). In particular, needle sharing was associated with younger age; users older than 45 years of age were less likely to share needles. Needle sharing was also associated with less education, current unemployment, and a history of having experienced homelessness, arrest, or imprisonment in the previous 10 years. The homosexual/bisexual male group had the highest proportion of individuals who had recently shared needles (81.6%). However, needle sharing was not statistically associated with either race or gender.

All variables examined in univariate analysis were entered into a logistic regression model to estimate the adjusted relative odds associated with that variable's contribution to needle sharing. After adjustment for all other variables, the only demographic variables that remained associated with needle sharing were younger age and education. Older and more educated individuals were less

TABLE 2—Life History Factors Associated with Needle Sharing during 6 Months Prior to Last Injection, Baltimore, Md, February 1988 through January 1990 (n = 2524)

	No. Participants	% Who Shared	Univariate OR (95% CI)	Adjusted OR (95% CI)[a]
Life events				
Public assistance last 10 y				
Yes	1765	73.3	1.55 (1.29,1.86)	1.41 (1.15,1.73)
No	759	63.9		
Arrested last 10 y				
Yes	1934	73.0	1.64 (1.35,1.99)	1.39 (1.13,1.72)
No	590	67.5		
Homeless last 10 y				
Yes	1227	73.5	1.33 (1.12,1.58)	1.09 (0.91,1.32)
No	1297	67.5		
Drug involvement				
Drug use				
Cocaine	2332	71.6	1.92 (1.42,2.59)	1.74 (1.27,2.39)
Other	192	56.8		
Frequency of injection				
≥ Daily	1008	77.1	1.73 (1.44,2.07)	1.44 (1.18,1.74)
Daily	1516	66.0		
Year started use				
1985 or before	2095	71.7	1.42 (1.14,1.77)	1.08 (0.85,1.37)
1986 or later	429	64.1		
Split drugs				
Yes	323	78.9	1.67 (1.26,2.21)	1.50 (1.12,2.01)
No	2201	69.2		
Experienced serious withdrawal				
Yes	717	80.5	2.08 (1.69,2.56)	1.73 (1.39,2.16)
No	1807	66.5		
Treatment last 10 y				
Ever	1326	74.4	1.49 (1.25,1.77)	1.20 (0.99,1.45)
Never	1198	66.1		

Note. OR = odds ratio; CI = confidence interval.
[a]From logistic regression model presented in text, which controlled for demographic, life event, and drug involvement variables.

likely to share needles. Among life-style variables, after adjustment for other variables, history of arrest in the previous 10 years, being on public assistance, and male–male sexual behavior remained significantly associated (i.e., the confidence interval excludes 1) with needle sharing. The variables describing current involvement with drugs (i.e., cocaine use, daily use, splitting drugs, and having experienced serious drug withdrawal symptoms) continued to demonstrate elevated odds for needle sharing after statistical adjustment. However, duration of injection drug use and history of treatment for drug abuse were not statistically associated with current needle sharing in the multivariate analysis.

Discussion

The primary mechanism for transmission of HIV infection among injecting drug users, needle sharing, has traditionally been viewed as a ritual firmly embedded within a subculture that is highly hesitant to change.[15] More recent study has suggested a pragmatic view that drug users also have an economic motivation to share needles.[10] The economic motivation is that drug users have limited funds and their priority is to obtain drugs. The findings of this study, which noted a higher proportion of needle sharing among the more economically disadvantaged drug users, are consistent with this pragmatic interpretation of needle sharing.

In addition to economic motivation, the finding of an association between needle sharing and history of involvement with the criminal justice system is consistent with a legal motivation to share needles. In particular, these data support findings of ethnographic studies that reported that drug users are hesitant to possess injection equipment that puts them at risk for arrest on charges of drug paraphernalia possession; prior experience with arrest and incarceration might serve as a deterrent to carrying injection equipment.[16,17] If so, the laws prohibiting possession of paraphernalia appear to have limited impact on illicit drug use per se, but they may have contributed to conditions that facilitate transmission of HIV infection. At the time paraphernalia laws were enacted, it is unlikely that this result was anticipated.

An alternate interpretation of the findings reported here is that persons who share needles are less socially competent and have fewer resources than those who do not share needles. Homelessness and being on public assistance can be considered as indicators of a seriously limited ability to function in society. Arrest is associated with lesser competence as a criminal. Such marginal individuals might be more dependent on others for access to drugs and injection equipment. The positive association of needle sharing and splitting drugs (a common practice among those who purchase drugs jointly with other users) could be consistent with this interpretation. Thus, needle sharing might to some degree be a marker of limited social competence. The public health implication is that resources are needed to engage this population.

The positive association between needle sharing and being a homosexual or bisexual injection drug user is disturbing. This finding suggests that this dual-risk group served as an important bridge for transmission of HIV infection to the injection drug–using community earlier in the epidemic.[18] Especially within minority communities, homosexual and bisexual men are relegated to lower status and therefore may be less able to control the conditions under which they use drugs. Although considerable efforts have been expended to establish prevention programs separately for homosexual men and injection drug users, additional special efforts are needed for homosexual and bisexual male injection drug users.

The positive association between needle sharing and higher frequency of injection and cocaine injection is indicative of the interrelatedness between intensity of drug use and risky injection practices. This interrelatedness underscores the importance of targeting interventions toward those with the highest levels of involvement with illicit drug use.

The findings of this study complement those of earlier studies that noted that needle sharing is frequent among injection drug users. Injection behavior is typically learned from more experienced

users.[19,20] A user's first drug injections are usually not planned[21] and typically require borrowing needles and taking instruction from a more experienced user who models the behavior. This initial joint use establishes early in the drug career a norm of sharing needles; continued needle sharing probably supports a social bond within injection drug–using groups. However, this study suggests that continued needle sharing is probably due also to pragmatic factors, including economic and legal motivations. Although the model for understanding needle sharing is likely to be complex, the results of this and other studies offer a basis for public health intervention. Recent studies of injection drug users show concern over personal health and intent to change high-risk practices[22,23]; trends in risk reduction have been observed.[24] Barriers to further change need to be addressed. Legal and economic barriers are reduced by needle exchange programs that offer unrestricted free and legal access to needles and syringes as part of a comprehensive HIV prevention program.[25] These data provide additional support for expanding needle exchange programs to provide a variety of outreach services to injection drug users. □

Acknowledgments

This research was supported by National Institute on Drug Abuse grants DA-04334 and DA-04341.

References

1. Chaisson RE, Moss AR, Onishi R, Osmond D, Carlson JR. Human immunodeficiency virus infection in heterosexual intravenous drug users in San Francisco. *Am J Public Health.* 1987;77:169–172.
2. Marmor M, Des Jarlais DC, Cohen H, et al. Risk factors for infection with human immunodeficiency virus among intravenous drug abusers in New York City. *AIDS.* 1987;1:39–44.
3. Schoenbaum EE, Hartel D, Selwyn PA, et al. Risk factors for human immunodeficiency virus infection in intravenous drug abusers. *N Engl J Med.* 1989;321:874–879.
4. Friedland GH, Harris C, Butkus-Small C, et al. Intravenous drug users and the acquired immunodeficiency syndrome: demographic, drug use, and needle sharing patterns. *Arch Intern Med.* 1985;145:141–147.
5. Turner CF, Miller HG, Moses LE, eds. *AIDS, Sexual Behavior, and Intravenous Drug Use.* Washington, DC: National Academy Press; 1989.
6. Brickner PW, Torres RA, Barnes M, et al. Recommendations for control and prevention of human immunodeficiency virus (HIV) infection in intravenous drug users. *Ann Intern Med.* 1989;110:833–837.
7. Celentano DD, Vlahov D, Menon AS, Polk BF. HIV knowledge and attitudes among IVDUs—comparisons to the US population and by drug use behaviors. *J Drug Issues.* 1991;21:647–661.
8. Vlahov D, Muñoz A, Celentano DD, et al. HIV seroconversion and disinfection of injection equipment among intravenous drug users, Baltimore, Maryland. *Epidemiology.* 1991;2:444–446.
9. Black JL, Dolan MP, DeFord HA, et al. Sharing of needles among users of IV drugs. *N Engl J Med.* 1986;314:446–447. Letter.
10. Magura S, Shapiro JL, Grossman JI, et al. Determinants of needle sharing among intravenous drug users. *Am J Public Health.* 1989;79:459–462.
11. Alcabes P, Vlahov D, Anthony JC. Correlates of human immunodeficiency virus infection in intravenous drug users: are treatment-program samples misleading? *Br J Addict.* 1992;87:47–54.
12. Vlahov D, Anthony JC, Muñoz A, et al. *The ALIVE Study, a Longitudinal Study of HIV-1 Infection in Intravenous Drug Users: Description of Methods and Characteristics of Participants.* Rockville, Md: National Institute on Drug Abuse; 1991:75–100. NIDA Research Monograph no. 109.
13. Anthony JC, Vlahov D, Celentano DD, et al. Self-report interview data for a study of HIV-1 infection among intravenous drug users: description of methods and preliminary evidence on validity. *J Drug Issues.* 1991;21:739–757.
14. Breslow NE, Day NE, eds. *Statistical Methods in Cancer Research, Vol. I: The Analysis of Case-Control Studies.* Lyon, France: International Agency for Research on Cancer; 1980. IARC scientific publication no. 32.
15. Des Jarlais DC, Friedman SR, Strug D. AIDS and needle sharing within the intravenous drug use subculture. In: Feldman D, Johnson T, eds. *The Social Dimensions of AIDS: Methods and Theory.* New York, NY: Praeger Publishers Inc; 1986.
16. Hanson B, Beschner G, Watters JM, Bowille E. *Life with Heroin: Voices from the Inner City.* Lexington, Mass: D.C. Heath; 1985.
17. Fiddle S. *Portraits from a Shooting Gallery.* New York, NY: Harper & Row; 1967.
18. Battjes RJ, Pickens RW, Amsel Z, Brown LS Jr. Heterosexual transmission of human immunodeficiency virus among intravenous drug users. *J Infect Dis.* 1990;162:1007–1011.
19. Powell DH. A pilot study of occasional heroin users. *Arch Gen Psychiatry.* 1973;28:586–594.
20. Harding W, Zinberg N. The effectiveness of the subculture in developing rituals and sanctions for controlled drug use. In: DuToit B, ed. *Drugs, Rituals and Altered States of Consciousness.* Rotterdam, the Netherlands: A.A. Balkena; 1977.
21. Des Jarlais DC, Friedman SR, Spira TJ, eds. A stage model of HTLV-III/LAV infection in intravenous drug users. In: *Problems of Drug Dependence.* Proceedings of the 47th Annual Scientific Meeting, Committee on Problems of Drug Dependence. Rockville, Md: National Institute on Drug Abuse; 1986:328–334. NIDA Research Monograph No. 67.
22. Selwyn PA, Feiner CP, Cox C, Lipshutz C, Cohen RL. Knowledge about AIDS and high risk behavior among intravenous drug users in New York City. *AIDS.* 1987;1:247–254.
23. Friedman SR, Des Jarlais DC, Sotheran JL. AIDS health education for intravenous drug users. *Health Educ Q.* 1986;13:383–393.
24. Vlahov D, Anthony JC, Celentano DD, Solomon L, Chowdhury N. Trends of risk reduction among initiates into intravenous drug use 1982–1987. *Am J Drug Alcohol Abuse.* 1990;17:39–48.
25. Gostin LO. The needle-borne HIV epidemic: causes and public health responses. *Behav Sci Law.* 1991;9:257–304.

The Effects of Early Education Intervention on Maternal Employment, Public Assistance, and Health Insurance: The Infant Health and Development Program

Jeanne Brooks-Gunn, PhD, Marie C. McCormick, MD, Sam Shapiro, BS, April Ann Benasich, PhD, and George W. Black, PhD

ABSTRACT

Objectives. The purpose of this study was to test whether early education intervention influences maternal employment, education, fertility, and receipt of public assistance and health insurance.

Methods. The Infant Health and Development Program is a randomized trial of the efficacy of early education on the outcomes of 985 low-birthweight, premature children. Families in eight sites received either pediatric follow-up and referral (follow-up only group) or pediatric services plus early intervention services (intervention group) for the first 3 years of the child's life.

Results. Mothers in the intervention group were employed more months and returned to the work force earlier than those in the follow-up only group. Fertility and education were not associated with treatment. Mothers who had some college education received more months of public assistance in the intervention group compared with the follow-up only group. Mothers who were employed received more public assistance and public health insurance in the intervention group compared with the follow-up only group, when maternal employment was controlled.

Conclusions. Findings are discussed in terms of the recent emphasis on two-generational programs directed to providing health, welfare, and child care services to young children and their families. (*Am J Public Health* 1994;84:924–931)

Introduction

Early childhood intervention programs were developed to help close the gap between disadvantaged children and their more advantaged peers by raising poor children's levels of social and educational competence in anticipation of elementary school. Since the middle 1960s, federal, state, community, and university-based programs have been initiated. The most carefully evaluated (with the exception of Head Start) are the small-scale programs conducted by educational researchers.[1–7]

The programs vary on a number of dimensions—delivery setting (home, school, center, clinic); primary target (mother, child, both); timing of onset (prenatal, infant, toddler); intensity (amount of programming per week); extensivity (length of program); curriculum content; ratios of staff to children; and staff training. However, all programs focus on enhancing child competence. To do so, programs employ a variety of strategies, including (1) working directly with the child, (2) helping the mother improve her interactions and teaching skills with her child, (3) teaching the mother about problem-solving abilities and child rearing, (4) raising the mother's self-esteem and emotional functioning, (5) promoting a return to school or the job market, and (6) procuring health and educational services for the child. All strategies but the first one assume that altering the mother's behavior or attitudes will influence the child indirectly. Indeed, many scholars believe that the long-term success of programs does not depend on their ability to directly alter the cognitive abilities of the child in the preschool years but on their capacity to alter the environment in which the child lives—characteristics of the mother; economic or residential circumstances of the mother; or health, welfare, and educational services offered to the child through the mother's efforts.[8–12]

This idea has widespread support.[4,13,14] However, not all studies have evaluated effects of programs on the mothers; those that have often do not report these effects in the literature, but rather in the final project reports.[15–19] Additionally, many of the early intervention studies were begun in the late 1960s and 1970s, before the rapid increase of mothers of young children in the work force.[20] The provision of full-time, high-quality, center-based child care, as offered by many early intervention programs, is of great benefit to mothers wishing to return to work, especially in the early years of the child's life, when high-quality child care is in such limited supply.[21–23]

This article focuses on the effects on the mothers of early intervention services for families with a low-birthweight, premature infant, specifically with regard to maternal employment, education, and fertility, as well as the use of public assistance (Aid to Families with Dependent Children) and publicly funded health services (Medicaid). The Infant Health

Jeanne Brooks-Gunn is with the Educational Testing Service, Princeton, NJ, and Columbia University, New York, NY. Marie C. McCormick is with the Harvard University School of Public Health, Boston, Mass. Sam Shapiro is with the Johns Hopkins University School of Public Health, Baltimore, Md. April Ann Benasich is with Rutgers University, New Brunswick, NJ. George W. Black is with Stanford University, Stanford, Calif.

Requests for reprints should be sent to Jeanne Brooks-Gunn, PhD, Center for Children and Families, Teachers College, Columbia University, New York, NY 10027.

This paper was accepted August 10, 1993.

and Development Program, a randomized trial, offered home visiting from the time of the birth of a low-birthweight, premature child to the child's third birthday and center-based child development programming in the second and third years of life. Home visitors provided information on child development, health and safety, social and medical services, and problem-solving skills. The last two might contribute to receipt of public assistance, health insurance, and employment. The provision of full-time, high-quality child care might benefit employment and education. It was expected that full-time employment benefits and education benefits would be found, with the effects being more likely after the children are 12 months of age, the age at which center-based early intervention began. Stronger maternal employment effects were expected for the less-educated mothers, who often are unable to afford quality child care.[20,22,24] The provision of early intervention services also might affect the number of mothers receiving public assistance and using public health insurance.[25–27] Reductions would be expected, in part because of the fact that more mothers would be in the work force. At the same time, the Infant Health and Development Program staff might have helped women receive services for which they were eligible, which would result in more women from the intervention group receiving services. Although the intervention did not target fertility (as no family planning or maternal postnatal care services were offered), it could indirectly affect fertility via increased health care use. The hypothesis that the intervention indirectly causes a reduction in fertility was tested by comparing the proportion of mothers in the two groups who had a birth during the follow-up period.

TABLE 1—Comparison of Intervention and Follow-Up Only Groups at Randomization[a]: The Infant Health and Development Program

	Intervention (n = 366)[d]	Follow-Up Only (n = 595)[d]
Child characteristics		
Mean (SD) birthweight,[b] g	1819 (439)	1781 (468)
Mean (SD) score on neonatal health index[c]	100.7 (16.0)	99.6 (15.8)
Percentage male[b]	50	49
Maternal characteristics[d]		
Mean (SD) education,[b] y	12.0 (2.6)	12.4 (2.4)
Mean (SD) age,[b] y	24.6 (5.9)	24.9 (6.1)
Ethnicity,[b] %		
Black	53	52
Hispanic	10	11
White/other	37	37
Percentage employed (at randomization)	31	30
Percentage employed (before pregnancy)	80	79
Percentage in school (at randomization)	11	12

[a]No significant differences between groups.
[b]Randomization procedure included birthweight, gender, maternal education, maternal age, and maternal ethnicity.
[c]Neonatal health index is length of neonatal stay controlled by birthweight and site, standardized so that the mean = 100 and the SD = 16; high scores indicate better health.
[d]Numbers may vary because of missing data for maternal characteristics.

Methods

Design

The Infant Health and Development Program is a multicenter, randomized, controlled trial to evaluate the efficacy of an intervention consisting of early child development programs and family support services in reducing the prevalence of health and developmental problems among low-birthweight, premature infants (in this study, defined as infants with a birthweight of ≤ 2500 g and a gestational age of ≤ 37 weeks). The intervention model and the specific curricula used for low-birthweight infants in the Infant Health and Development Program were adapted from two longitudinal studies of early intervention with full-term children of low-income families: the Abecedarian Project and Project CARE.[28,29] Comprehensive child and parent curricula were delivered through home visiting, parent groups, and children's attendance at a child development center program. The two treatment groups were an intervention group that received all three of the above curricula plus high-quality pediatric follow-up through age 3 years and a follow-up only group that received only the pediatric follow-up. Referrals were made as necessary. Sites were responsible for assisting families, when needed, to obtain access to health care; three sites chose to provide primary care for their subjects.

The intervention program was initiated immediately after randomization on the infant's discharge from the hospital and continued until a corrected age of 3 years (i.e., the age was corrected for the estimated number of weeks premature). The protocol specified four home visits per month in year 1 and two visits per month in years 2 and 3. On average across all sites, families received about three visits per month in year 1 and 1.5 visits per month in years 2 and 3.[30] The child development centers provided the intervention group children with an enriched, extrafamilial, educational experience using a learning-game format beginning at 12 months (corrected age) and continuing until the last child at the site reached 36 months of age (corrected age). The children were scheduled to attend the center at least 4 hours daily, 5 days per week. The mean number of hours attended per day was 5. The comprehensive educational program emphasized cognitive, linguistic, and social development.[31]

The primary outcomes in the Infant Health and Development Program were related to the child's cognitive development and behavioral competence, areas in which low-birthweight premature infants are at risk for adverse outcomes. Secondary outcomes focused on mothers. The provision of the program's services may affect the families by allowing mothers to take advantage of educational or employment opportunities and to have increased access to and use of health care and community resources. Analyses on access to and use of health care services are reported in McCormick et al.[32]

Sample

Informed consent was sought from the parents of 1442 eligible infants at eight sites (sites are listed in Acknowledgements). Omitting one twin from each family and refusals, 985 infants were designated as the primary analysis group and constitute the cohort for the trial (see Table 1; for a description of the sample, see reference 33).

Infants were randomized to either the intervention group or the follow-up

only group immediately after hospital discharge, by use of a procedure that involved close monitoring for balance and for absence of bias.[34,35] The characteristics for which balance was sought were birthweight, gender, maternal age, maternal education, maternal race (Black, Hispanic, and White/other), primary language in the home, and infant participation in another study.[33,36,37]

Assessment Schedule

Children in both the intervention and follow-up only groups received pediatric surveillance, as offered through eight clinic visits at 40 weeks of gestational age and at 4, 8, 12, 18, 24, 30, and 36 months postterm age. These visits also were used to collect assessment data. All assessments were conducted in English.

Measures

Maternal outcomes. Outcomes include maternal employment, education, fertility, and receipt of public health insurance and public assistance. Mothers indicated whether they were employed, in school, or receiving public assistance at each clinic visit (4, 8, 12, 18, 24, 30, and 36 months postterm age). These data were used to calculate the cumulative months employed, in school, and receiving public assistance, as well as the timing of work force entry (focusing on the first entry mentioned, not on repeated entrances and exits). Mothers were asked whether they used public health insurance or had any health insurance at 36 months. The number of subsequent births was determined by the responses at the 12-, 24-, and 36-month assessments. Given the skewed nature of the distributions, fertility was coded as having had an additional child by the 36-month assessment (yes/no).

Initial status measures. Several initial status variables were examined because of associations with the primary outcome measures in previous research. Demographic factors include ethnicity, maternal education, and maternal age in years. Sites were examined because they differed on demographic factors.[33,38] Child characteristics were gender, birthweight, and score on a neonatal health index (a measure that standardizes the length of the neonatal hospital stay for birthweight and is converted into a scale with a mean of 100 and a SD of 16, with higher scores indicating better health).[39]

Data Analyses

Multiple linear regressions were performed to test for treatment group effects, controlling for the possible effects of the initial status variables just mentioned. Site was entered as seven dummy codes, and each site was compared with the eighth site. Maternal ethnicity was dummy-coded into two variables (Black vs Hispanic and White/other; Hispanic vs Black and White/other). Maternal education was coded as three categories (less than high school, high school graduation, and some college), maternal age in years, birthweight in grams, and neonatal health on a continuous scale. Additionally, possible interactions of the initial status variables with treatment group were examined in these regressions for all outcome measures.[33]

To see whether group membership influenced the timing of entry into the work force, we conducted Kaplan-Meier survival curve analyses. Additionally, we conducted subgroup analyses using multiple linear regressions to examine effects by ethnicity and education simultaneously. Mothers were divided into those with some college and those with a high school education or less. Effects for Black and White mothers were examined separately. Separate analyses for Hispanics with a high school education or less also were conducted. For these analyses, only sites with at least five families in the intervention group and five in the follow-up only group were included (see reference 38 for a description of the rationale for this procedure and descriptive statistics). Four sites were included for the group that was White and had some college (Arkansas, Harvard, Washington, Yale); four for the group that was Black and had some college (Arkansas, Einstein, Harvard, Pennsylvania); six for the group that was White and had a high school education (Arkansas, Einstein, Harvard, Texas, Washington, Yale); all eight for the group that was Black and had a high school education or less; and three for the group that was Hispanic and had a high school education or less (Einstein, Miami, and Texas).

Results

Maternal Employment

Cumulative months of employment. Mothers in the intervention group were employed for a greater number of months than were mothers in the follow-up only group (1.84 months, $P = .04$). This effect was net of site, demographic, and child characteristics, as found in the multiple linear regression analysis. Maternal age and race did not contribute to the cumulative number of months worked, although maternal education did (5.73 months per category of education, $P = .001$). The vast majority of the mothers in this trial were employed: 80% of the intervention group mothers and 72% of the Follow-up only group mothers worked at least 4 months out of the 3 years of the trial. At the same time, 16% of the intervention group mothers and 21% of the follow-up only mothers worked the entire 36 months.

Although no main effects of child characteristics were found in the multiple linear regression, a significant interaction between treatment and birthweight was found ($P = .02$). Investigation of the mean number of months that mothers worked for the two birthweight groups revealed that the treatment group differences appeared in the lighter, not the heavier, birthweight group. Mothers in the intervention group who had lighter infants (weighing 2000 g or less at birth) worked more months than mothers in the follow-up only group with lighter infants (mean of 18.1 months vs 15.3 months). Mothers of heavier infants (in the 2001- to 2500-g group) worked 14.3 months in the intervention group and 16.2 months in the follow-up only group.

Subgroup analyses indicated that the beneficial effect of the intervention on maternal employment occurred in mothers with a high school degree or less (Table 2). Black mothers were employed an average of 2.51 months more in the intervention group compared with the follow-up only group ($P = .05$); White mothers, an average of 5.05 months more in the intervention group compared with the follow-up only group ($P = .02$). (These differences are slightly different than those presented in Table 2, because they are based on the regression analyses controlling for initial characteristics, whereas the data in Table 2 are simple means.)

Timing of entry into the work force. The Kaplan-Meier survival analysis for timing of first reported entry into the work force by child's age (corrected for prematurity) is presented by treatment group in Figure 1. As can be seen, mothers in the intervention group were more likely to enter the work force somewhat earlier than the mothers in the follow-up only group ($P < .06$). The differences appeared when the children were 18 months of age.

Given the interaction between treatment and birthweight found in the analysis of cumulative months of employment, we examined the timing of the first

TABLE 2—Cumulative Months of Maternal Employment, Education, and Public Assistance, by Treatment Group, Maternal Education, and Ethnicity: Infant Health and Development Program

	Mean (SD) Employment, mo		Mean (SD) Education, mo		Mean (SD) Public Assistance, mo	
	INT (n = 300)[b]	FU (n = 486)[b]	INT (n = 308)[b]	FU (n = 500)[b]	INT (n = 307)[b]	FU (n = 500)[b]
Total sample	16.7 (12.9)	15.6* (14.2)	4.9 (9.0)	4.2 (2.6)	14.4 (15.8)	12.6 (15.2)
College education						
Black[a]	23.9 (11.2)	25.5 (13.0)	7.6 (12.4)	3.8 (7.6)	12.0 (16.1)	5.4** (11.1)
White[a]	22.4 (14.2)	21.8 (14.6)	2.4 (7.6)	4.1 (7.9)	3.2 (10.4)	0.2* (1.6)
High school education or less						
Black	14.1 (12.1)	12.0* (12.8)	6.2 (9.3)	5.0 (8.0)	20.6 (15.1)	20.3 (14.9)
White	18.1 (11.5)	14.0* (13.2)	2.3 (5.8)	2.9 (6.0)	8.4 (13.1)	9.5 (13.5)
Hispanic	7.7 (11.3)	9.9 (12.6)	7.5 (10.8)	3.5 (7.7)	22.8 (16.5)	20.4 (16.6)

Note. INT = intervention group; FU = follow-up only group.
[a]Means for education by ethnicity subgroup include mothers from those sites that had five or more families in both the intervention and follow-up only groups (see reference 36 and the text). Too few Hispanic mothers had some college to analyze.
[b]Numbers vary somewhat because of missing data (measures are aggregated over time)
*$P < .05$; **$P < .10$.

work-force experience reported after the child's birth for the lighter and heavier low-birthweight infants. For the infants with birthweights of 2000 g or less, mothers in the intervention group were likely to enter the work force earlier than those in the follow-up only group, with differences first emerging when the child was about 8 months of age ($P < .001$). In contrast, no significant differences were found for the heavier infants.

Cox proportional hazards analyses were used to test differences in timing of entry into the work force. Variables entered into the regressions were treatment, site, maternal characteristics (education, age, ethnicity), and child characteristics (gender, neonatal health status, birthweight). A significant treatment effect was found ($P = .05$), as mothers in the intervention group entered the work force earlier than mothers in the follow-up only group. Mothers with more education also entered the work force earlier than mothers with less education ($P = .0001$).

Maternal Education

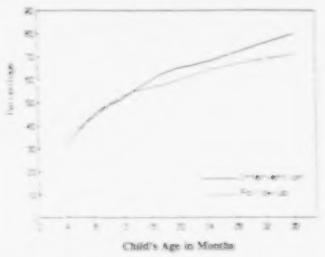
Multiple linear regression analyses on the cumulative number of months in school over the 3 years of the Infant Health and Development Program were conducted, paralleling the analyses for maternal employment. No significant treatment effect was found (Table 2).

The multiple linear regressions suggest that younger mothers spent more months in school than older mothers (0.33 months per year of age; $P = .001$), and Black mothers reported more months of schooling than other mothers (1.90 months, $P = .01$; see also Table 2). An interaction between treatment and race ($P = .04$) suggested that Hispanic mothers in the Intervention group received more months of education than those in the follow-up only group (4.8 months), whereas that was not true of the other two groups of mothers. No significant treatment effects were seen in the subgroup analyses.

Fertility

About one third (36%) of the total group of mothers had at least one additional birth in the 3 years since the target child's arrival (whether a subsequent birth had occurred was used in the regression rather than number of births because of the skewed nature of the distribution; numbers were 304 for the intervention group and 501 for follow-up only group). No treatment effect was found (2% fewer births in the treatment group; $P = .48$).

Not surprisingly, the multiple linear regressions suggest that younger mothers were more likely to have had a subsequent birth (2% of the women had an additional birth per year of age; $P = .0001$). Controlling for all other variables, Black and Hispanic mothers were somewhat more likely to have a subsequent birth (7% and 12%, respectively; $P = .08$ and $P = .06$, respectively). No interactions with treatment group were found, nor were any of the subgroup analyses significant. (To have an 80% chance of detecting a difference between the two groups (two-sided $\alpha = .05$), the absolute difference in proportions would have to be .10 or larger. The difference here was .02.)

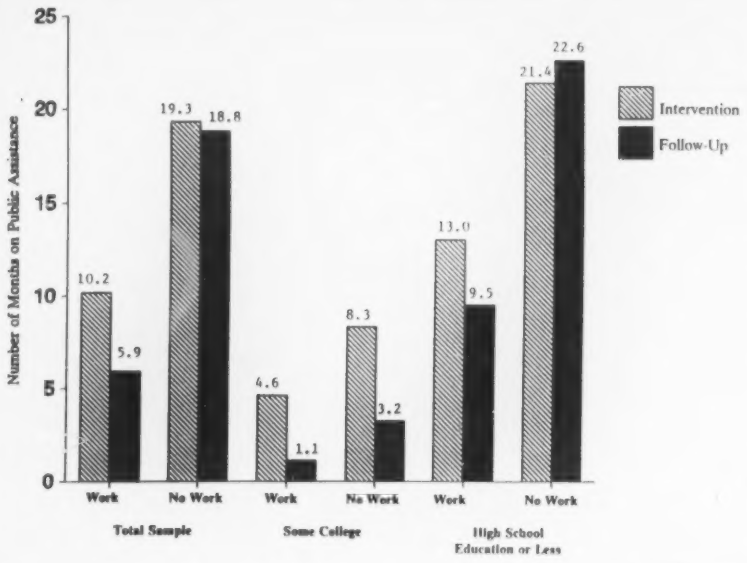


Note. Kaplan-Meier survival analysis. Child's age is postterm.

FIGURE 1—Timing of entry into the work force for mothers in the intervention and follow-up groups.

Public Assistance

Receipt of public assistance during the 36 months was not associated with treatment group. Large effects were found for ethnicity and education ($P = .001$ for each; see mean values in Table 2). However, a significant interaction between treatment and maternal education ($P = .005$) suggests that the more-educated mothers in the intervention group were more likely to receive public assis-



Note. Sample restricted to sites with five or more children in both the intervention and follow-up groups (see Brooks-Gunn et al.[38]). No Hispanics are included in the group with some college. Work groups: n = 395 for mother not employed (No Work); n = 402 for mother employed when the child was 36 months old (Work). Data are simple means based on subgroup analyses (maternal employment status when the child was 36 months old, treatment group, and maternal education when the child was born).

FIGURE 2—Number of months that mothers received public assistance, by treatment group, maternal employment status at 36 months, and maternal education.

tance than more-educated mothers in the follow-up only group. This was true for the Black subgroup (6.27 months more public assistance; $P = .07$) and the White subgroup (2.42 months more public assistance; $P = .02$).

In a separate regression analysis where maternal employment at 36 months was included as an independent variable, a significant treatment effect was seen ($P = 0.03$). That is, when controlling for maternal employment (which is highly negatively associated with the receipt of public assistance, as expected; $P = .005$), mothers in the intervention group were more likely to receive public assistance (1.78 months longer). A trend toward an interaction between treatment and employment ($P = .09$) suggests that mothers who were employed received more months of public assistance if they were in the intervention group compared with the follow-up only group (Figure 2). A significant interaction between maternal education and treatment group ($P = .04$) suggests that treatment group mothers who had some college education received more months of public assistance, but no association with treatment was found for mothers who had a high school education or less (Figure 2).

Public Medical Insurance

About 40% of the sample reported receiving public health insurance at the 36-month assessment (300 in the intervention group and 472 in the follow-up only group). Such receipt was not associated with treatment group. Overall, 82% of the mothers reported that they had some form of health insurance. Maternal ethnicity, education, and age were not associated with receipt of health insurance. No treatment effects for any form of health insurance were seen, either in the initial analyses or in the second set of multiple linear regressions when maternal employment was entered as a dependent variable. A significant interaction between maternal education and treatment group ($P = .05$) indicates that mothers with more education were more likely to have public health insurance when they were in the intervention group compared with the follow-up only group.

When maternal employment at 36 months is entered into the regression as a independent (control) variable, the intervention group was more likely to have public health insurance (the difference between the two groups was 7%; $P = .02$). A trend ($P = .08$) for an interaction between maternal employment and treatment group suggests that the treatment effects were found for the mothers who were employed at the 36-month point (Figure 3). An interaction between ethnicity and treatment group indicates that the effects were most pronounced for the Hispanic mothers ($P = .03$).

Discussion

Like other early childhood interventions in the first 3 years of life, the Infant Health and Development Program has the potential to influence mothers as well as children. This program differs from most previous studies of interventions that continue past the first years of life by focusing on a biologically vulnerable group—low-birthweight infants—rather than an environmentally vulnerable group—disadvantaged children.[2,40] However, about two thirds of the families in the Infant Health and Development Program were disadvantaged, as defined by low maternal education levels, and income. Children from these families are at increased risk for subsequent problems[10,41] and their mothers may be as well.[32] The inclusion of Hispanic and White families also is a welcome and significant departure from most of the early intervention literature, which focused primarily on Black families.[42,43]

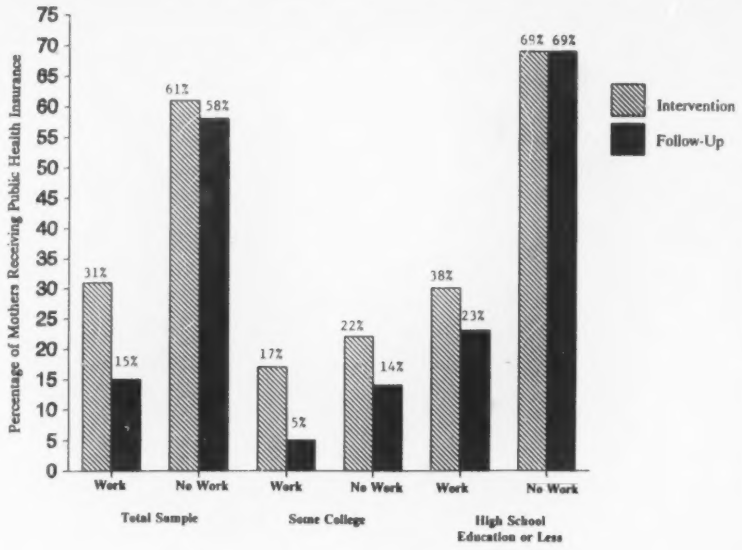
Results from the Infant Health and Development Program parallel those of previous studies with disadvantaged children in that the provision of early intervention services for children under the age of 3 years was found to increase the likelihood of maternal employment. As in other studies, these effects were modest but significant. Benasich et al. reviewed 27 programs providing educationally based intervention services in the first 3 years of life.[15] Of the 11 programs that reported on maternal employment as an outcome, all 5 home-based programs and 5 of the 6 center-based programs found that mothers in the programs were more likely to be employed or to be more stably employed than mothers in the comparison groups. Studies tend to report relatively small effects (e.g., treatment effects lasting less than 4 months). With respect to the timing of entrance into the work force, the

fact that group differences began to appear when children were 18 months of age suggests that the effect was primarily due to the provision of center-based care, which started when the child was between 12 and 13 months of age.[44]

Of importance is the fact that the maternal employment effects were found for mothers with a high school education or less. Other studies have not included mothers with college education. Mothers with less education (who earn less money) are probably unable to purchase high-quality child care,[20] making the Infant Health and Development Program very attractive.

Earlier studies did not examine timing of entry into the work force. Timing is particularly critical given the current debates about the effects of maternal employment and center-based care on children in the first year of life.[20,44,45] In the Infant Health and Development Program about half of the mothers of low-birthweight infants had entered the work force by the infant's first birthday, which is similar to the proportion reported in national samples of mothers.[20] When low-birthweight and normal-birthweight infants were compared by using the data for children in the National Longitudinal Study of Youth,[46] mothers with low-birthweight infants were somewhat less likely to place their children in outside-the-home care in the first year of life than were mothers with normal-birthweight infants.[47] Still, half of these mothers entered the work force during the first year of their child's life. In the Infant Health and Development Program, although actual birthweight was not associated with number of months worked, mothers of infants who had lower birthweights tended to enter the work force somewhat later than mothers of infants who weighed more ($P = .08$).

Of interest is the fact that being in the intervention group (as opposed to the follow-up only group) was associated with a longer period of employment for mothers with lighter low-birthweight infants but not for mothers with heavier low-birthweight infants. Generally, these results may indicate that mothers of lighter low-birthweight infants are reluctant to leave them for long periods of time or to put them in organized child care because the lighter children are sicker or are perceived to be sicker. Indeed, mothers in the Infant Health and Development Program did report that the lighter low-birthweight children experienced more hospital care (hospitalizations and number of hospital days), ambulatory care, and illnesses than the heavier low-birthweight children.[32,33] Thus, the family services provided to the mothers in the intervention group may have lessened their concerns about the health of the lighter low-birthweight infants, either through the information given in the home visits or the availability of high-quality, center-based care with excellent pediatric surveillance.



Note. Sample restricted to sites with 5 or more children in both the intervention and follow-up groups (see Brooks-Gunn et al.[38]). No Hispanics are included in the some-college group. Work groups: n = 382 for mother not employed (No Work); n = 390 for mother employed when the child was 36 months old (Work). Data are percentages based on subgroup analyses (maternal employment status when the child was 36 months old, treatment group, and maternal education when the child was born).

FIGURE 3—Percentages of mothers receiving public health insurance when the child was 36 months old, by treatment group, maternal employment status at 36 months, and maternal education.

Treatment effects were not found for schooling in the Infant Health and Development Program, unlike some of the previous interventions with disadvantaged children.[15] The intervention group mothers also did not differ from the follow-up only group mothers with respect to subsequent fertility. Mothers in the intervention group did not have lower fertility rates, but they also did not have higher fertility rates (a possibility, given the provision of free child care). Few earlier educational programs examined fertility, in part because family planning services were not offered in these programs. Only 15% of the 27 studies reviewed by Benasich et al.[15] included data on fertility. All reported positive effects and were comprehensive programs,[48–52] and two focused on teenage mothers and their children.[53,54] However, it is possible that longer-term fertility may be affected, rather than fertility at the completion of the program during the target child's toddler-hood years.[53–55]

Receipt of public assistance and health insurance were not examined as maternal outcomes in most of the early childhood intervention literature. However, our findings indicate that treatment effects were seen when maternal employment was controlled for. Of interest is the fact that mothers who were employed were more likely to receive these services when they were in the treatment group, suggesting that these mothers were eligible for public assistance and that the Infant Health and Development Program staff informed them of eligibility and procedures. These mothers presumably were employed in jobs that were low wage, part time, or "off the books." The ability of an early educational intervention program to improve access to health care via health insurance is relevant from a policy perspective, given that the number of families without health insurance

has increased.[26,56,57] Use of health care services was more frequent in the treatment group of the Infant Health and Development Program. These differences appeared at 12 months (results not shown), suggesting that the home-visiting component of the program played a role in the receipt of these services (because center-based schooling did not begin until 12 months).

A limitation of the study is the use of maternal report. Virtually all national studies rely on mothers to indicate employment, education, and fertility history. Reports of public assistance are fairly accurate (when compared with welfare records) over a much longer time period than that used in this study.[55] Additionally, the mothers in the Infant Health and Development Program were asked about these events every 4 to 6 months, which is a shorter recall period than that used in most studies.

Almost all early intervention programs postulate indirect or mediating effects of maternal and parenting outcomes on the child.[58] However, most employ a child-focused evaluation and only look at direct effects on the child. Indirect effects may constitute the "transmission pathways" for long-term effects on the child.[4,13,14] These include not only parenting behavior but also the ability to procure health, child care, and welfare services, as illustrated here. Additionally, enhancing maternal functioning is an important outcome in its own right.[9] Finally, more emphasis is being placed on two-generational approaches to providing services to poor young children and their families[59–61] in the hopes that both will benefit and that good maternal outcomes and service receipt will enhance the likelihood that the children will be positively affected.

The Family Support Act of 1988 is an excellent example of a focus not only on two generations but on integrated services. To help mothers leave the welfare system and enter the work force, they are offered education and job training as well as transitional child care (through a voucher system) and health insurance (through Medicaid). Both of the latter are critical in that the jobs for unskilled or semiskilled workers, such as young mothers with a high school degree or less, pay too little to purchase child care services and usually do not provide health insurance.[61] The results from the Infant Health and Development Program reported here illustrate that providing services to families—even though the focus is on the child and parenting behavior—has the potential to increase employment and to increase the receipt of health and welfare services for those not employed. □

Acknowledgments

The Infant Health and Development Program was funded by the Robert Wood Johnson Foundation. Additional support was provided from the Pew Charitable Trusts; the Bureau of Maternal and Child Health and Resources Development, Health Resources Service Administration; the National Institute of Child Health and Human Development; and the March of Dimes Birth Defects Foundation. Their generosity is appreciated.

The participating universities and site directors were Patrick H. Casey, MD, University of Arkansas for Medical Sciences (Little Rock, Ark); Cecelia M. McCarton, MD, Albert Einstein College of Medicine (Bronx, NY); Michael W. Yogman, MD, Harvard Medical School (Boston, Mass); Charles R. Bauer, MD, and Keith G. Scott, PhD, University of Miami School of Medicine (Miami, Fla); Judy Bernbaum, MD, University of Pennsylvania School of Medicine (Philadelphia, Pa); Jon E. Tyson, MD, and Mark Swanson, MD, University of Texas Health Science Center at Dallas; Clifford J. Sells, MD, and Forrest C. Bennett, MD, University of Washington School of Medicine (Seattle, Wash); and David T. Scott, PhD, Yale University School of Medicine (New Haven, Conn).

We thank Ruth Gross, Director of the National Study Office, and Craig Ramey, Director of the Program Development Office, Infant Health and Development Program, for their help. We also thank Rosemary Deibler for her assistance in manuscript preparation.

References

1. Andrews SR, Blumenthal JB, Johnson DL, et al. The skills of mothering: a study of Parent-child development centers. *Monogr Soc Res Child Dev.* 1982;47(6). Serial No. 198.
2. Brooks-Gunn J. Promoting healthy development in young children: what educational interventions work? In: Ginzburg E, Rogers DE, eds. *Improving the Chances of Children at Risk.* Boulder, Colo: Westview Press; 1990;125–145.
3. Haskins R. Beyond metaphor. *Am Psychol.* 1989;44:274–282.
4. Lazar I, Darlington R, Murray H, Royce J, Snipper A. Lasting effects of early education: a report from the Consortium for Longitudinal Studies. *Monogr Soc Res Child Dev.* 1982;47(2–3). Serial No. 195.
5. McKey RH, Condelli L, Granson H, Barret B, McConkey C, Plantz M. *The Impact of Head Start on Children, Families and Communities.* Final report of the Head Start Evaluation, Synthesis and Utilization Project. Washington, DC: CSR, Inc; 1985.
6. Zigler EF, Muenchow S. How to influence social policy affecting children and families. *Am Psychol.* 1984;39:415–420.
7. Zigler EF, Valentine J. *Project Head Start: A Legacy of the War on Poverty.* New York: Macmillan; 1979.
8. Bronfenbrenner U. Contexts of child rearing: problems and prospects. *Am Psychol.* 1979;34:844–850.
9. Olds DL. Can home visitation improve the health of women and children at risk? In: Ginzburg E, Rogers DE, eds. *Improving the Chances of Chilren at Risk.* Boulder, Colo: Westview Press; 1990:79–103.
10. McCormick MC, Brooks-Gunn J. Health care for children and adolescents. In: Freeman H, Levine S, eds. *Handbook of Medical Sociology.* Englewood Cliffs, NJ: Prentice Hall; 1989:347–380.
11. Meisels SJ, Shonkoff JP. *Handbook of Early Childhood Intervention.* Cambridge, United Kingdom: Cambridge University Press; 1990.
12. Woodhead M. When psychology informs public policy: the case of early childhood intervention. *Am Psychol.* 1988;43:443–454.
13. Clement JRB, Schweinhart LJ, Barnett WS, Epstein AS, Weikart DP. *Changed Lives: The Effects of the Perry Preschool Program on Youths through Age 19.* Ypsilanti, Mich: High/Scope; 1984. Monograph No. 8.
14. Gray SW, Ramsey BK, Klaus RA. *From 3 to 20: The Early Training Project.* Baltimore, Md: University Park Press; 1982:311–362.
15. Benasich AA, Brooks-Gunn J, Clewell BC. How do mothers benefit from early intervention programs? *J Appl Dev Psychol.* 1992;13:311–362.
16. Clarke-Stewart KA, Fein GG. Early childhood programs. In: Mussen PH, ed. *Handbook of Child Psychology.* 4th ed. New York, NY: John Wiley & Sons; 1983:918–999.
17. Gray SW, Wandersman LP. The methodology of home-based intervention studies: problems and promising strategies. *Child Dev.* 1980;51:993–1009.
18. Ramey CT, Bryant DM, Suarez TM. Preschool compensatory education and the modificability of intelligence: a critical review. In: Detterman D, ed. *Current Topics in Human Intelligence.* Norwood, NJ: Ablex Publishing Corp; 1985:247–296.
19. White K, Casto G. An integrative review of early intervention efficacy studies with at-risk children: implications for the handicapped. *Analysis and Intervention in Developmental Disabilities.* 1985;5:7–31.
20. Hayes CD, Palmer JL, Zaslow MJ. *Who Cares for America's Children: Child Care Policy for the 1990's.* Washington, DC: National Academy Press; 1990.
21. Liebowitz A, Waite L, Witsberger C. Child care for preschoolers: differences by child's age. *Demography.* 1988;25:205–220.
22. Hofferth SL, Phillips D. Child care in the United States, 1970–1995. *J Marrige Fam.* 1987;49:559–571.
23. Phillips DA. *Quality in Child Care: What Does Research Tell Us?* Washington, DC: National Association for the Education of Young Children; 1987.
24. Waite LJ, Leibowitz A, Witsberger C. What parents pay for: child care characteristics, quality, and costs. *J Soc Issues.* 1991;47:33–48.
25. Egbuonu L, Starfield B. Child health and social status. *Pediatrics.* 1982;69:550–557.
26. Newacheck PW. Improving access to care for expectant mothers and young children. In: Ginzburg E, Rogers DE, eds. *Improving the Chances of Children at Risk.* Boulder, Colo: Westview Press; 1990:63–78.

27. Newacheck PW, McManus MA. Financing health care for disabled children. *Pediatrics.* 1988;81:385–394.
28. Wasik BH, Ramey CT, Bryant DM, Sparling JJ. A longitudinal study of two early intervention strategies: Project CARE. *Child Dev.* 1990;61:1682–1696.
29. Ramey CT, Campbell FA. The Carolina Abecedarian Project: an educational experiment concerning human malleability. In: Gallagher JJ, Ramey CT, eds. *The Malleability of Children.* Baltimore, Md: Brookes Publishing Co; 1987:127–140.
30. Ramey CP, Bryant DM, Wasik BH, Sparling JJ, Fendt KH, LaVange LM. The Infant Health and Development Program for low birthweight, premature infants: program elements, family participation, and child intelligence. *Pediatrics.* 1992;3: 454–465.
31. Sparling JJ, Lewis IS, Neuwirth S. *Early Partners* [curriculum kit]. Lewisville, NC: Kaplan Press; 1991.
32. McCormick MC, Brooks-Gunn J, Shapiro S, Benasich AA, Black G, Gross RT. Health care use among young children in day care: results seen in a randomized trial of early intervention. *JAMA.* 1991;265:2212–2217.
33. Infant Health and Development Program Staff. Enhancing the outcomes of low-birthweight, premature infants: a multisite randomized trial. *JAMA.* 1990;263:3035–3042.
34. Efron B. Forcing a sequential experiment to be balanced. *Biometrika.* 1971;58:403–417.
35. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics.* 1975;31:103–115.
36. Brooks-Gunn J, Klebanov PK, Liaw F, Spiker D. Enhancing the development of low birth weight, premature infants: changes in cognition and behavior over the first three years. *Child Dev.* 1993;64:736–753.
37. Kraemer HC, Fendt KH. Random assignment in clinical trials: issues in planning. *J Clin Epidemiol.* 1990;43:1157–1167.
38. Brooks-Gunn J, Gross RT, Kraemer HC, Spiker D, Shapiro S. Enhancing the cognitive outcomes of low-birth-weight, premature infants: for whom is the intervention most effective? *Pediatrics.* 1992;89:1209–1215.
39. Scott DT, Bauer CR, Kraemer HC, Tyson J. Neonatal health index for preterm infants. *Pediatr Res.* 1989;25:263a.
40. Bennett FC. The effectiveness of early intervention for infants at increased biological risk. In: Guralnick M, Bennett FC, eds. *The Effectiveness of Early Intervention for At-Risk and Handicapped Children.* New York, NY: Academic Press; 1987:79–112.
41. Parker S, Greer S, Zuckerman B. Double jeopardy: the impact of poverty on early child development. *Pediatr Clin North Am.* 1988;35:1227–1240.
42. Bridgeman B, Blumenthal JB, Andrews RR. *Parent Child Development Centers: Final Evaluation Report.* Submitted to the Department of Health and Human Services, Washington, DC, 1981.
43. Lee V, Brooks-Gunn J, Schnur E. Does Head Start "close the gap"? A comparison of children attending Head Start, no preschool, and other preschool programs. *Dev Psychol.* 1988;24:210–222.
44. Baydar N, Brooks-Gunn J. Effects of maternal employment and child-care arrangements in infancy on preschoolers' cognitive and behavioral outcomes: evidence from the children of the NLSY. *Dev Psychol.* 1991;27:932–945.
45. Desai S, Chase-Lansdale L, Michael RT. Mother or market? Effects of maternal employment on the intellectual ability of 4-year-old children. *Demography.* 1990;26: 545–561.
46. Chase-Lansdale L, Mott F, Brooks-Gunn J, Phillips D. Children of the NLSY: a unique research opportunity. *Dev Psychol.* 1991;27:918–931.
47. Mott FL. Developmental effects of infant care: the mediating role of gender and health. *J Soc Issues.* 1991;47:139–158.
48. Gordon IJ, Guinagh BJ. *A Home Learning Center Approach to Early Stimulation.* Final report to the National Institute of Mental Health. In: *JSAS Catalog of Selected Documents in Psychology.* Gainesville, Fla: Institute for Development of Human Resources, University of Florida; 1974:8. Manuscript No. 1634.
49. Olds DL, Henderson CR, Tatelbaum R, Chamberlin R. Improving the delivery of prenatal care and outcomes of pregnancy: a randomized trial of nurse home visitation. *Pediatrics.* 1986;77:16–28.
50. Olds DL, Henderson CR, Tatelbaum R, Chamberlin R. Preventing child abuse and neglect: a randomized trial of nurse home visitation. *Pediatrics.* 1986;78:16–28.
51. Olds DL, Henderson CR, Tatelbaum R, Chamberlin R. Improving the life-course development of disadvantaged mothers: a randomized trial of nurse home visitation. *Am J Public Health.* 1988;78:1436–1445.
52. Seitz V, Rosenbaum LK, Apfel NH. Effects of family support intervention: a ten year follow-up. *Child Dev.* 1985;56:376–391.
53. Field T, Widmayer S, Greenberg R, Stoller S. Effects of parent training on teenage mothers and their infants. *Pediatrics.* 1982; 69:703–707.
54. Roosa MW, Vaughn L. Teen mothers enrolled in an alternative parenting program: a comparison with their peers. *Urban Education.* 1983;18:348–360.
55. Furstenberg FF Jr, Brooks-Gunn J, Morgan P. *Adolescent Mothers in Later Life.* New York, NY: Cambridge University Press; 1987.
56. Enthoven A, Kronick R. A consumer-choice health plan for the 1990s. *N Engl J Med.* 1989;320:29–37.
57. Rosenbach ML. The impact of medicaid on physician use by low-income children. *Am J Public Health.* 1989;79:1220–1226.
58. Bronfenbrenner U. Is early intervention effective? In: Guttentag M, Streuning E, eds. *Handbook of Evaluation Research.* Beverly Hills, Calif: Sage; 1975;2:519–603.
59. Duncan G, Brooks-Gunn J, Klebanov PK. Economic deprivation and early-childhood development. *Child Dev.* In press.
60. Ginzberg E, Rogers DE, eds. *Improving the Chances of Children at Risk.* Boulder, Colo: Westview Press; 1990.
61. Chase-Lansdale PL, Brooks-Gunn J. *Escape from Poverty: What Makes a Difference for Children.* New York, NY: Cambridge University Press. In press.

The Size of Mortality Differences Associated with Educational Level in Nine Industrialized Countries

Anton E. Kunst, MA, and Johan P. Mackenbach, MD, PhD

ABSTRACT

Objectives. This study addresses the question of whether inequalities in premature mortality related to educational level differ among countries.

Methods. Data on mortality by educational level were obtained from longitudinal studies from nine industrialized countries. The data referred to men between 35 and 64 years of age. The follow-up periods occurred between 1970 and 1982. The size of mortality differences associated with educational level was measured by means of two inequality indices, both based on Poisson regression analysis.

Results. Inequalities in mortality are relatively small in the Netherlands, Sweden, Denmark, and Norway and about two times as large in the United States, France, and Italy. Finland and England and Wales occupy intermediate positions. The large inequalities in mortality in the United States and France can be attributed in part to large inequalities in education in these countries.

Conclusions. The international pattern found in this study was also observed in a comparison that used occupation as the socioeconomic indicator. Differences between countries in levels of inequality in mortality may be partially explained by the countries' different levels of egalitarian social and economic policies. (*Am J Public Health.* 1994;84:932–937)

Introduction

Inequalities in mortality that are related to socioeconomic status are a generalized phenomenon in the industrialized world. In each country for which data are available, death rates have been found to be higher in groups with lower occupational status, educational level, or income level.

The question of whether these inequalities in mortality are about equally large in all countries or whether important differences exist has aroused wide interest. The primary reason for this interest is that inequalities in other countries provide a point of reference for judging whether health inequalities in one's own country are small or large. In addition, cross-country comparisons provide a unique opportunity to assess the effects of national socioeconomic policies on the magnitude of health inequalities.

Cross-country comparisons have traditionally used occupational status as the socioeconomic indicator.[1–8] Most of these studies were confronted with serious data problems, such as international differences in occupational classifications and a lack of information on the occupational class of large sections of the population (women, old men, middle-aged economically inactive men).[9] There is therefore a strong need to complement these studies with international comparisons that are based on other indicators of socioeconomic status.

The aim of the present study was to compare a large number of countries with respect to mortality differences associated with educational level. This study refers to men in the age group 35 to 64 years. Data were obtained from studies done in the United States and a number of European countries.

Methods

Data on mortality by educational group were acquired for as many countries as possible. Studies were selected that met the following criteria: (1) the study design was longitudinal, (2) the observation period occurred in the 1970s or the early 1980s, and (3) the study population was nationally representative. The selected studies[3,10–17] are presented in Table 1. The US study[17] consisted of a follow-up of eight samples of the Current Population Survey. The Dutch study[10] was an epidemiological follow-up study of military conscripts born in 1930. All other European studies selected were follow-ups of the national population census, either for a representative sample (England and Wales[14] and France[15]) or for the entire national population.

The length of follow-up in most studies was about 10 years. Studies with shorter follow-up times were also included because from analyses of data from the Netherlands, Finland, and France we found that the size of educational mortality differences does not change with increasing follow-up length.

Table 2 presents the educational classifications used in the respective stud-

The authors are with the Department of Public Health, Faculty of Medicine and Health Sciences, Erasmus University, Rotterdam, the Netherlands.

Correspondence should be sent to Anton Kunst, MA, Department of Public Health, Faculty of Medicine and Health Sciences, Erasmus University, PO Box 1738, 3000 DR Rotterdam, the Netherlands.

This paper was accepted February 16, 1994.

ies. The average number of years of education was estimated by the authors in consultation with contact persons from the respective countries.

The educational level was not given for more than 1% of the populations of Denmark, Sweden, and England and Wales. In each of these cases, it was assumed that the educational level of this "unknown" group was equal to the level of the group with the lowest educational level. Support for this assumption in the Swedish case is that the occupational composition of the "unknown" group was similar to that of the group in the lowest educational category (Ö. Hemström and D. Vågerö, PhD, written communication, February 26, 1992). For Denmark, a comparison with more complete data from the Danish Health and Morbidity Survey suggested that the group with "unknown" education in the mortality study should have a low educational level. Because the mortality level of the "unknown" group in England and Wales was close to the national average, alternative assumptions about the group's educational level had a negligible effect on the measurement of the association between mortality and educational level.

Two other data problems are briefly mentioned here and will be discussed more extensively at the end of the Results section. First, the educational level of men in the Dutch study was assessed when they were 18 or 19 years old, that is, when some of them had still not left school. Second, more than two thirds of the population in Sweden, Finland, and England and Wales was assigned to a single (the lowest) educational category.

We compare countries by means of inequality indices that, for each country separately, measure the size of mortality differences associated with educational level. Indices were selected on the basis of two criteria: (1) all educational groups should be included separately in the calculation of the index, and (2) not all mortality differences should be measured, but only mortality differences that are systematically related to an ordering of educational groups from high to low socioeconomic status.[2]

These two criteria can be met by means of regression analysis. We applied Poisson (log-linear) regression, using the following regression model:

$$D_{are} = P_{are} \cdot e^{(\alpha_{ar} + \beta SES_e)},$$

TABLE 1—Studies Included in the International Comparison

Country	Reference	Follow-Up Period	Age Groups Distinguished, y[a]	Sample Size[b]
Netherlands	Doornbos & Kromhout[10]	1970–1981	38	79 000
Denmark	Andersen[11,c]	1970–1980	35–39, 40–44	289 000
Norway	Valkonen[12]	1971–1980	30–49[d]	141 000
Sweden	Vägerö & Lundberg[3,c]	1971–1980	35–39, 40–44, . . . 55–59	1 204 000
Finland	Valkonen[13,c]	1970–1980	35–39, 40–44, . . . 60–64	732 000
England & Wales	Fox & Goldblatt[14,c]	1971–1981	15–44, 45–64	162 000
France	Desplanques[15]	1975–1980	35–44, 45–54, 55–64	314 000
Italy	Pagnanelli[16]	1981–1982	18–54[d]	10 468 000
United States	Rogot et al.[17]	1979–1981	35–44, 45–54, 55–64	130 000

[a]Age at beginning of follow-up, except for England and Wales and France, where data refer to age at death.
[b]Number of male subjects at the start of follow-up in the given age groups.
[c]Data were obtained from unpublished tables.
[d]Mortality figures presented as age-standardized death rates or standardized mortality ratios.

where D is the estimated number of deaths, P is the number of person-years at risk, and SES is a measure of the socioeconomic status of each educational group. The subscripts a, r, and e denote 5-year age group, race (United States only), and educational group, and α and β are regression coefficients. The formula $(\exp\beta - 1)$ yields the inequality index used in this study. It represents the proportional increase in mortality per one unit increase in socioeconomic status.

The socioeconomic status of educational groups can be quantified in two ways. The first way is to quantify the educational level as the average number of years that are formally needed to complete that level. The corresponding inequality index represents the proportional mortality increase associated with 1 additional year's education. This index has been applied in an earlier international comparison.[12]

In the second method, the socioeconomic status of an educational group is conceptualized as the group's relative position in the social hierarchy. Following Pamuk,[18] this position is quantified as the proportion of the population that has a higher position in the social hierarchy. For example, if the highest educational group comprises 10% of the population, the relative position of its members would be between 0 and 0.10, the average being 0.05. The corresponding inequality index represents the proportional mortality increase associated with one unit increase in the relative position. More specifically, the index can be interpreted as the proportional mortality increase by moving from the top (= 0) to the bottom (= 1) of the social hierarchy.

We call the first index the partial inequality index and the second the total inequality index. These names are related to their interpretation, which can be illustrated as follows. If the score of a country on the total index is large compared with the scores of other countries, this implies large mortality differences between high and low positions on the social hierarchy. This large mortality difference can be attributed to large differences between high and low social positions in the number of years of education (i.e., large educational inequalities) or a large effect of 1 extra year of education on mortality (i.e., a large score on the partial inequality index). Thus, while the partial index measures the size of the effect of 1 extra year of education on mortality, the total index also takes into account the extent of inequality in educational levels themselves. The total inequality index therefore measures the total size of mortality differences in a population that are related to educational inequality. Since the two measures are complementary and each has a specific interpretation, both will be applied in the present study.

The regression equation assumes that mortality has a log-linear relationship with educational level. This assumption

TABLE 2—Educational Classifications Applied in the Various Studies

Country	Educational Groups	Average No. of Years of Education	% of Total Male Study Population[a]
Netherlands[10,b]	Primary education	8.0	48.2
	Lower vocational education	10.0	21.0
	Lower secondary education	12.0	17.8
	Higher level	15.5	12.9
Denmark[11]	9 years' education or less	8.0	27.8
	Educational level unknown	8.0	12.6
	10–12 years' education	11.0	40.0
	Vocational education, level unknown	11.5	7.1
	13–14 years' education	13.5	6.2
	15 years' education or more	16.0	6.3
Norway[12]	Primary school or not reported	7.5	38.9
	Second level, first stage	9.0	20.8
	Second level, second stage	11.0	27.6
	Third level, first stage	14.0	8.2
	Third level, second stage	17.0	4.4
Sweden[3]	Educational level unknown	8.0	2.6
	Primary education	8.0	81.6
	Some education after primary school	9.5	6.3
	Secondary education	14.0	9.5
Finland[13]	Basic education, lower level	8.0	76.6
	Basic education, upper level	9.0	3.1
	Secondary education, lower level	10.5	8.7
	Secondary education, upper level	12.0	5.8
	Higher education, lowest level	13.5	2.1
	Higher education, undergraduate level	15.0	0.9
	Higher education, postgraduate level	17.0	2.8
England and Wales[14]	Educational level unknown	10.0	2.2
	No higher qualifications	10.0	82.1
	A-level only	12.0	6.3
	Nondegree higher qualifications	14.0	5.0
	Degree or equivalent	17.0	4.4
France[15,c]	No qualifications	7.0	31.6
	Second level, first stage	10.0	46.1
	Second level, second stage	12.0	8.2
	Baccalaureate	15.0	6.6
	Higher education	18.0	7.6
Italy[16,d,e]	Elementary education not completed	...	3.9
	Elementary education completed	...	40.1
	Lower secondary education	...	31.5
	Upper secondary education	...	17.3
	Postsecondary education	...	7.2
United States[17]	Elementary, grades 1–4 or lower	3.0	2.8
	Elementary, grades 5–7	6.0	5.5
	Elementary, grade 8	8.0	8.2
	High school, years 1–3	10.0	14.8
	High school, year 4	12.0	35.3
	College, years 1–3	14.0	13.7
	College, year 4	16.0	10.0
	College, 5 years or more	18.0	9.7

[a]At the beginning of follow-up.
[b]Educational level was assessed when the subjects were 18 or 19 years old.
[c]Percentages in the last column refer to the entire French population instead of the study population because the latter is based on a sample stratified by socioeconomic group.
[d]No reliable estimates are available for number of years of education corresponding to the educational groups used in the Italian study.
[e]Estimated from data on the same cohort in the survey Health Conditions of the Population and the Use of Health Services, 1986–1987.

was checked by means of (1) inspection of residuals and (2) the inclusion of a quadratic term for the education measure in the regression equation. No large departures from (log) linearity were observed.

Results

Total Inequality Index

Table 3 presents measurements of the total amount of inequality in mortality. The size of inequality decreases strongly with increasing age. Since the precise delimitation of age groups differs among studies, comparisons between studies cannot readily be made on the basis of Table 3. To facilitate comparisons, inequality estimates are presented in Figure 1 according to mean age at death.

At age 55 years and older, the smallest inequalities in mortality were observed for Sweden. The value of 0.36 for the oldest age group implies, according to the fitted regression equation, that death rates estimated for those at the bottom of the Swedish educational hierarchy are 36% higher than the death rates estimated for those at the top. Larger inequalities in mortality are observed for, in order of magnitude, England and Wales, Finland, the United States, and France. The total amount of inequality in France is almost four times as large as that for Sweden.

At ages below 55 years, the differences between Sweden, Finland, and France follow the same pattern but are less pronounced. The total amount of inequality in mortality is less than two times as large in France as in Sweden. The small inequality estimate for England and Wales and the very large estimate for the United States might have resulted from large chance fluctuations.

Data on inequalities in mortality below the age of 55 years were available for four additional countries. The inequality estimate for the Netherlands is considerably smaller than that for Sweden, while the estimates for Norway and Denmark are nearly identical to that for Sweden. The inequality estimates for Italy and France are equally large.

Partial Inequality Index

In Figure 2, countries are compared with respect to the partial inequality index, that is, the proportional increase in mortality associated with 1 additional year of education. Italy could not be included because of the absence of reliable data on the number of years of education.

The rank order of countries from small to large partial inequality estimates is the same as the order shown in Figure 1

for the total inequality estimate. The only exception is the United States, for which the total amount of inequality in mortality (total index) is large but the effect of 1 additional year of education (partial index) is relatively small. Among the European countries, the rank order of countries is the same for both inequality indices, but the differences in the total amount of inequality in mortality (total index) are much larger than the differences in the effect of 1 additional year of education on mortality (partial index).

As explained in the Methods section, large mortality differences related to educational level (the total index) can be attributed to a large effect of educational level on mortality (the partial index) or to large inequalities in educational levels. The results suggest that the former explanation seems not to hold for the United States and to hold only to some extent for France. This suggests that the large total amount of inequality in the United States and France is, at least in part, related to large inequalities in educational levels themselves.

To document this inference, we used a measure of inequality in educational levels that could be applied to the data for most countries: the average number of years of education of the top 20% of the population minus the average number of years of education of the bottom 80% of the population. According to this approximate measure, inequalities in education in the United States (5.75 years) and France (6.75 years) are indeed larger than those in other countries (e.g., England and Wales [3.75 years] and Sweden [4.25 years]).

TABLE 3—The Size of Mortality Differences Associated with Educational Level among Men, by Country and Age Group

Country	Total Inequality Estimate[a] (95% Confidence Interval)		
	Age 35–44 y[b]	45–54 y	55–64 y
Netherlands	0.72 (0.51, 0.95)[c]	. . .	. . .
Denmark	1.17 (1.02, 1.33)	. . .	. . .
Norway	1.02 (0.89, 1.16)[d]	. . .	. . .
Sweden	1.20 (1.02, 1.41)	0.60 (0.51, 0.69)	0.36 (0.28, 0.45)[e]
Finland	1.49 (1.32, 1.68)	0.99 (0.90, 1.08)	0.79 (0.71, 0.87)
England and Wales[f]	1.04 (0.43, 1.92)	0.75 (0.49, 1.06)	. . .
France[g]	1.97 (1.60, 2.39)	1.59 (1.31, 1.90)	1.28 (1.02, 1.56)
Italy	1.85 (1.75, 1.97)[h]	. . .	. . .
United States	2.62 (1.26, 4.78)	1.06 (0.51, 1.80)	1.05 (0.69, 1.48)

[a]The total inequality estimate represents the proportional mortality increase moving from the top to the bottom of the educational hierarchy.
[b]Age at beginning of follow-up.
[c]The men in the Dutch study were 38 years old.
[d]The men in the Norwegian study were 30 to 49 years old.
[e]The oldest age group in the Swedish study was 55 to 59 years.
[f]The age groups studied in England and Wales were approximately 10 to 40 years and 40 to 60 years.
[g]The age groups studied in France were approximately 32 to 41 years, 42 to 51 years, and 52 to 61 years.
[h]The men in the Italian study were 18 to 54 years old.

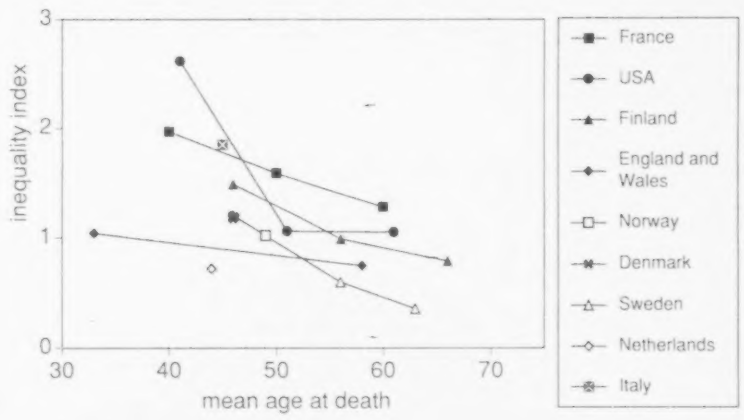


FIGURE 1—Total inequality index, by country and age group.

Evaluation of Potential Sources of Bias

Two potential sources of bias were briefly mentioned in the Methods section.

First, the educational level of subjects in the Dutch study was determined in 1950, when they were 18 or 19 years old. The level of education that an 18- or 19-year-old man had completed in 1950 was probably closely related to the educational level that he finally would achieve. However, discrepancies may exist, which implies misclassification of subjects. The most likely effect of such misclassification on inequality estimates is underestimation. Perhaps the underestimation is so large that it explains the relatively small observed inequalities for the Netherlands. We therefore feel that there is insufficient evidence to conclude that inequalities are smaller in the Netherlands than in Scandinavian countries.

Second, more than two thirds of the population of Sweden, Finland, and England and Wales was assigned to a single (the lowest) educational category. As a consequence, the relationship between mortality and educational level within a large segment of the population could not be included in the calculation of the inequality indices. Therefore, the inequality estimates are less accurate for these countries than for other countries. Perhaps, however, this problem is less serious than it seems. Recall that the relationship between mortality and educational level is approximately linear. This implies that approximately the same regression estimates would be obtained whether 4 or, say, 10 educational groups are distinguished. We found support for this expectation in additional analyses on all countries with detailed educational classifications. We reestimated inequality indices after combining the lower 80% of the population into one large educational category. Despite this substantial loss of information, the new inequality estimates

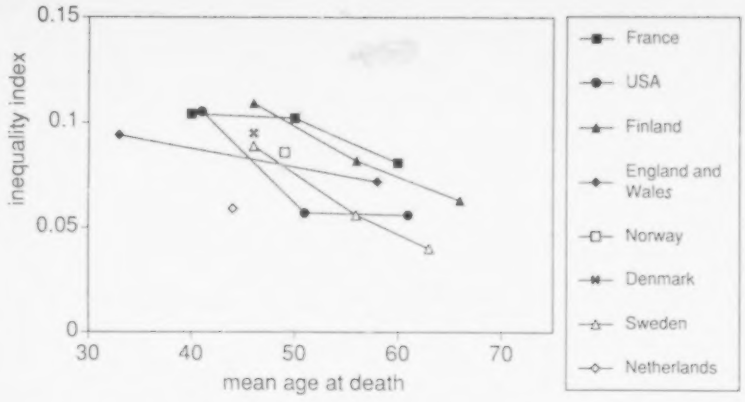


FIGURE 2—Partial inequality index, by country and age group.

were nearly the same as the estimates that were based on detailed educational classifications (less than 10% difference).

Discussion

An important question for the evaluation of socioeconomic inequalities in health is the extent to which these inequalities can be regarded as variable. This question has often been approached by time-series analysis, which has shown for various countries an increase in inequalities in mortality during the postwar period.[12,15,18–22] We chose a different approach by addressing the question, To what extent do inequalities in mortality vary across countries? Substantial cross-country differences were observed in the size of mortality differences associated with educational level. Differences are relatively small in the Netherlands, Sweden, Denmark, and Norway and about two times as large in the United States, France, and Italy. Finland and England and Wales occupy intermediate positions.

In addition, we investigated to what extent the large mortality differentials in the United States and France can be attributed to relatively large inequalities in educational levels themselves or to a large effect of educational level on mortality. We found that international variations in the size of inequalities in educational levels are most important, whereas the effect of education on mortality is approximately equally large in all countries.

The latter finding corroborates the fascinating finding of Valkonen's comparative study of England and Wales, Hungary, and the Scandinavian countries.[12] Valkonen found that in the 1970s, at ages below 50 years the relationship between education and mortality was "surprisingly similar": in each country death rates diminished by about 8% with an increase of 1 year in educational attainment. He warned, however, that this finding should not be taken as evidence of a "universal law" of health inequalities, because large cross-country differences were observed for specific causes of death and, in addition, might be observed for other time periods or other age groups. Indeed, we found that at ages older than 50 years, the effect of 1 year of additional education on mortality varied substantially among countries, with the effect in France being two times as large as that in Sweden.

We used education as the socioeconomic indicator because this indicator has been judged to be more comparable internationally than occupation.[9] Although in individual populations education and occupational status are strongly correlated,[23] the question remains whether the same international pattern of inequalities in mortality would be found if occupation were used as the socioeconomic indicator. The answer is probably affirmative: a recent study comparing seven European countries with respect to occupational mortality found virtually the same rank order of countries from small to large inequalities: (1) the Netherlands, Denmark, Norway, and Sweden; (2) England and Wales; (3) Finland; and (4) France.[24]

Cross-country comparisons provide a new opportunity to identify circumstances that are associated with large or small mortality differentials. A relevant question, therefore, is why the extent of inequalities in premature mortality varies among countries.

The first line of explanation relates to the selection hypothesis, which states that health and education are related in part because educational achievement depends on, among other things, health or health-related factors. The contribution of health selection to the generation of inequalities in adult mortality has been much disputed but is bound to be the subject of conjecture as long as lifelong longitudinal studies are not carried out.[25,26] Important for the present study is that the magnitude of health selection is likely to vary among countries as a function of educational and training structures.[26] Lahelma and Valkonen posited a more specific hypothesis: In countries where people coming from different strata have more equal access to education, the achieved educational level may depend less on social background and more on personal characteristics, including health and health-related factors. Paradoxically, then, health and education may be relatively closely related in open, competitive societies.[27] Testing this hypothesis requires detailed assessment of international variation in educational and training structures, which is outside the scope of this paper. The hypothesis may partially account for the fact that the United States, where intergenerational social mobility is larger than in nearly any other industrialized country,[28] is also one of the countries with the largest inequalities in both mortality and morbidity.

The alternative line of explanation relates to the causation hypothesis, which stresses the effect of educational level on health. The higher death rates of lower-educational-level groups are explained at least in part by a higher prevalence of risk factors for disease, such as factors related to life-styles, material living conditions, working conditions, and ways of coping with stress. A logical extension of this assertion is that international variations in the social distribution of risk factors explain, at least in part, the international variation in inequalities in mortality.

An illustration is offered by France and Italy. Leclerc et al. showed that diseases related to alcoholism and alcohol abuse accounted for a substantial part of the excess mortality of lower social classes in France, whereas their role is much more limited in Finland and in England and Wales.[1,29] Italian data suggest that to an important extent, the large mortality differences in Italy are also attributable to

excessive alcohol-related mortality among members of lower socioeconomic groups.[16]

Explanations cannot be confined to risk factors for disease, but should also consider the more distal social, economic, and cultural factors. One potentially relevant factor is welfare and income policies. Wilkinson's work is particularly relevant here.[30,31] He demonstrated that national death rates are strongly related to the extent of income inequality, and he inferred from this finding that large income inequalities are associated with large inequalities in mortality. Our study allows for a first test of this hypothesis. During the 1970s, income inequalities were relatively large in the United States, France, and Italy and approximately equally small in the other countries.[32,33] Income inequalities in Finland, still large in the 1960s, diminished rapidly during the 1970s. This rank order of countries in terms of income inequalities strongly corresponds to their rank order in terms of inequalities in mortality. Thus, our material corroborates Wilkinson's inferences. A main challenge to future international comparisons is to determine whether this association is spurious or reflects a true positive effect of egalitarian social and economic policies. □

Acknowledgments

This study would not have been possible without the unpublished data supplied by the following persons: Otto Andersen (Danish data), Denny Vågerö (Swedish data), Tapani Valkonen (Finnish data), and John Fox and Haroulla Filakti (data on England and Wales). Recent publications on France were supplied by Guy Desplanques. We thank Tapani Valkonen, Annette Leclerc, John Fox, Haroulla Filakti, and Denny Vågerö for their detailed comments on an earlier version of this paper.

References

1. Leclerc A, Lert F, Fabien C. Differential mortality: some comparisons between England and Wales, Finland and France, based on inequalities measures. *Int J Epidemiol.* 1990;4:1–10.
2. Wagstaff A, Paci P, Doorslaer E van. On the measurement of inequalities in health. *Soc Sci Med.* 1991;33:545–557.
3. Vågerö D, Lundberg O. Health inequalities in Britain and Sweden. *Lancet.* 1989;ii: 35–36.
4. Minder CE. A conceptual and methodological discussion of transnational comparisons of socio-economic mortality differences, and an example. In: *Proceedings of the Fifth Meeting of the UN/WHO/CICRED Network on Socio-Economic Differential Mortality in Industrialized Countries.* Paris, France: Committee for International Cooperation in National Research on Demography (CICRED); 1991:278–294.
5. Neumann G von, Liedermann A. Mortalität und Sozialschicht. *Bundesgesundheitsblatt.* 1981;24:173–181.
6. Ramis-Juan O, Sokou K. Social health inequalities in South European countries: is it a different problem? In: Fox J, ed. *Health Inequalities in European Countries.* Aldershot, England: Gower; 1989:199–230.
7. Kagaminori S, Tibuchi Y, Fox J. A comparison of socioeconomic differences in mortality between Japan and England and Wales. *World Health Stat Q.* 1983;36: 119–128.
8. Guralnick L. Mortality by occupational level and cause of death among men 20–64 years of age, United States, 1950. *Vital Stat Special Rep.* 1963;53:452.
9. Valkonen T. Problems in the measurement and international comparisons of socio-economic differences in mortality. *Soc Sci Med.* 1993;36:409–418.
10. Doornbos G, Kromhout D. Educational level and mortality in a 32-year follow-up study of 18-year-old men in the Netherlands. *Int J Epidemiol.* 1990;19:374–379.
11. Andersen O. *Dodelighed ag Erhverv 1970–1980.* Copenhagen, Denmark: Danmarks Statistics; 1985.
12. Valkonen T. Adult mortality and level of education: a comparison of six countries. In: Fox J, ed. *Health Inequalities in European Countries.* Aldershot, England: Gower; 1989:142–172.
13. Valkonen T. *Socio-Economic Mortality Differences in Finland 1971–1985.* Helsinki, Finland: Central Statistical Office of Finland; 1990.
14. Fox AJ, Goldblatt PO. *Longitudinal Study. Socio-Demographic Mortality Differentials.* London, England: Her Majesty's Stationery Office; 1982. Series LS no. 1.
15. Desplanques G. *La Mortalité des Adults: Résultats de 2 Etudes Longitudinales (Période 1955–1980).* Paris, France: Institut National d'Etudes Démographiques (INSEE); 1984.
16. Pagnanelli F. Une synthese des informations sur les différences de mortalité en Italie selon certains facteurs socio-économiques. In: *Proceedings of the Fifth Meeting of the UNDP/WHO/CICRED Network on Socio-Economic Differential Mortality in Industrialized Countries.* Paris, France: Committee for International Cooperation in National Research on Demography (CICRED); 1991.
17. Rogot E, Sorlie PD, Johnson NJ. *A Mortality Study of One Million Persons by Demographic, Social, and Economic Factors: 1979–1981.* Washington, DC: US Dept of Health and Human Services; 1988. DHHS publication NIH 88-2896.
18. Pamuk ER. Social class and inequality in mortality from 1921 to 1972 in England and Wales. *Popul Stud.* 1985;39:17–31.
19. Pappas G, Queen S, Hadden W, Fisher G. The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *N Engl J Med.* 1993;329:103–109.
20. Valkonen T, Matelin T, Rimpelä A, Notkola V, Savela S. *Socio-Economic Mortality Differences in Finland 1981–90.* Helsinki, Finland: Statistics Finland; 1993.
21. Dahl E, Kjaersgaard P. Trends in socioeconomic mortality differentials in post-war Norway: evidence and interpretations. *Sociol Health Illness.* 1993;15:587–611.
22. Kunst AE, Looman CWN, Mackenbach JP. Socio-economic mortality differences in the Netherlands in 1950–1984: a regional study of cause-specific mortality. *Soc Sci Med.* 1990;31:141–152.
23. Ganzeboom HBG, de Graaf PM, Treiman DJ. A standard international socio-economic index of occupational status. *Soc Sci Res.* 1992;21:1–56.
24. Kunst AE, Mackenbach JP. International variation in the size of mortality differences associated with occupational status. *Int J Epidemiol.* In press.
25. Dahl E, Kjaersgaard P. Social mobility and inequality in mortality. An assessment of the health selection hypothesis. *Eur J Public Health.* 1993;3:124–132.
26. West P. Rethinking the health selection explanation for health inequalities. *Soc Sci Med.* 1991;32:373–384.
27. Lahelma E, Valkonen T. Health and social inequities in Finland and elsewhere. *Soc Sci Med.* 1990;27 (special issue):257–265.
28. Ganzeboom HBG, Luijkx R, Treiman DJ. Intergenerational class mobility in comparative perspective. *Res Social Stratification Mobility.* 1989;8:3–84.
29. Leclerc A, Lert F, Goldberg M. Les inegalites sociales devant la mort en Grande-Bretagne et en France. *Soc Sci Med.* 1984;19:479–487.
30. Wilkinson RG. Income distribution and life expectancy. *BMJ.* 1992;304:165–168.
31. Wilkinson RG. National mortality rates: the impact of inequality? *Am J Public Health.* 1992;82:1082–1084.
32. Taylor CL, Jodice DA. *World Handbook of Political and Social Indicators. Vol. I: Cross-National Attributes and Rates of Change.* New Haven, Conn: Yale University Press; 1983.
33. *National Accounts Statistics: Compendium of Income Distribution Statistics.* New York, NY: United Nations, Department of International Economic and Social Affairs; 1985. Statistical Papers, series M, no. 79.

Why Did Black Life Expectancy Decline from 1984 through 1989 in the United States?

ABSTRACT

Objectives. The objective of this study was to partition the change in US life expectancy into those major causes of death and age groups that contributed most to the decline in life expectancy for Black males and females and to the increase in life expectancy for White males and females in the period from 1984 through 1989.

Methods. By means of a life table partitioning technique, the positive and negative contributions of age and cause of death to changes in life expectancy were estimated.

Results. Causes contributing most to the decrease in life expectancy for Black males included human immunodeficiency virus (HIV) infection (<5 and 20–69 years of age) and homicide (1–4 and 10–34 years of age); for Black females the causes included HIV infection (<5 and 20–59 years of age) and cancer (65–79 and ≥85 years of age).

Conclusions. Mortality trends between the two major race groups in the United States diverged sharply, and without precedent, during the late 1980s. A description of these trends in terms of contributions to change in life expectancy may help us better analyze and interpret changes in the health of these groups and may contribute to a better allocation of resources for research, education, and public health programs. (*Am J Public Health.* 1994;84:938–944)

Kenneth D. Kochanek, MA, Jeffrey D. Maurer, MS, and Harry M. Rosenberg, PhD

Introduction

Since 1900, life expectancy at birth in the United States has generally increased for males and females and for both major race groups—the White and Black populations. However, from 1984 through 1989 a countervailing trend drew wide public attention: whereas life expectancy generally increased for both White males and females, it decreased for both Black males and females[1] (Table 1).

In this research, we partition the changes in life expectancy at birth into those major causes of death and age groups that contributed most substantially to the decline for Black males and females and to the increase for White males and females. Also examined separately are the changes in the "gap," or difference, in the life expectancy between Black and White males and between Black and White females for the same time period.

The analysis uses a life table disaggregation technique developed by Eduardo E. Arriaga[2,3] to "measure the pace of mortality change . . . as well as the decomposition of a difference in life expectancy."[2(p95)] A summary measure like life expectancy is a convenient measure of the mortality of a population. Life expectancy has the advantage of ready interpretation.[4]

Methods

To measure changes in mortality, Arriaga[2,3] developed a method to estimate the contributions of age and cause of death to changes in life expectancy, which he describes as a procedure that "estimates the number of years added to or removed from life expectation because of the decrease or increase (respectively) of the central mortality rates of life tables."[3] This method partitions the changes in life expectancy over time or between two separate groups of populations. In this report Arriaga's technique is used to partition changes in life expectancy at birth in the United States by age and cause of death from 1984 through 1989 for the total population and for the four major race-sex groups. These changes in life expectancy by race, sex, and age are partitioned into component additive parts.

The causes of death used in this study are based on the 15 leading causes of death in 1984 and 1989 for the total population and for each of the four major race-sex groups, resulting in a total of 17 causes. All other causes are in a residual category. The codes and complete titles of the causes of death consistent with the *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death*, Ninth Revision (ICD-9)[5] are shown in the Appendix. Cause-of-death rankings are based on the National Center for Health Statistics' "List of 72 Selected Causes of Death"[6] and the category of human immunodeficiency virus (HIV) infection (category numbers *042–*044), which was added to the lists of rankable causes effective with data year 1987. Further information on cause-of-death lists and rankable causes can be obtained from the Technical Appendix

The authors are with the Mortality Statistics Branch, Division of Vital Statistics, National Center for Health Statistics, Hyattsville, Md.

Requests for reprints should be sent to Kenneth D. Kochanek, MA, National Center for Health Statistics, 6525 Belcrest Rd, Room 840, Hyattsville, MD 20782.

This paper was accepted September 10, 1993.

section of *Vital Statistics of the United States.*[6] The number of deaths from HIV infection in 1984 was estimated by using the number of deaths reported with mention of ICD-9 code 279.1 (deficiency of cell-mediated immunity).

Because the changes in life expectancy, expressed in years, are relatively small when partitioned into age groups and causes of death, life expectancies were recomputed and are presented to three significant digits after the decimal.

TABLE 1—Life Expectancy (in Years) at Birth, by Race and Sex: United States, 1970 through 1989

		White			Black		
Year	Total Population	Both Sexes	Males	Females	Both Sexes	Males	Females
1989	75.3	76.0	72.7	79.2	69.2	64.8	73.5
1988	74.9	75.6	72.3	78.9	69.2	64.9	73.4
1987	75.0	75.6	72.2	78.9	69.4	65.2	73.6
1986	74.8	75.4	72.0	78.8	69.4	65.2	73.5
1985	74.7	75.3	71.9	78.7	69.5	65.3	73.5
1984	74.7	75.3	71.8	78.7	69.7	65.6	73.7
1983	74.6	75.2	71.7	78.7	69.6	65.4	73.6
1982	74.5	75.1	71.5	78.7	69.4	65.1	73.7
1981	74.2	74.8	71.1	78.4	68.9	64.5	73.2
1980	73.7	74.4	70.7	78.1	68.1	63.8	72.5
1979	73.9	74.6	70.8	78.4	68.5	64.0	72.9
1978	73.5	74.1	70.4	78.0	68.1	63.7	72.4
1977	73.3	74.0	70.2	77.9	67.7	63.4	72.0
1976	72.9	73.6	69.9	77.5	67.2	62.9	71.6
1975	72.6	73.4	69.5	77.3	66.8	62.4	71.3
1974	72.0	72.8	69.0	76.7	66.0	61.7	70.3
1973	71.4	72.2	68.5	76.1	65.0	60.9	69.3
1972	71.2	72.0	68.3	75.9	64.7	60.4	69.1
1971	71.1	72.0	68.3	75.8	64.6	60.5	68.9
1970	70.8	71.7	68.0	75.6	64.1	60.0	68.3

Note. Data are rounded to the nearest 0.1 year as published.[1]

Results

By Age

From 1984 through 1989, life expectancy at birth increased for the total population by 0.6 year (Table 1).[1] During the same period, life expectancy at birth increased for both White males and White females (by 0.9 and 0.5 year, respectively) but decreased for both Black males and Black females (−0.8 and −0.2 year, respectively).

Changes in life expectancy over time can be viewed as the net effect of "positive contributions" and "negative contributions" to life expectancy. Positive contributions result from declining death rates specific for age, race, sex, and cause of death, whereas negative contributions result from increasing death rates specific for these categories. Positive contributions from declines in one cause of death can be entirely offset by negative contributions from another cause for the same population group.

From 1984 through 1989, for the total population (males and females of all races combined), life expectancy at birth increased by about half a year, or 0.513 year (Table 2, "By Age"), as a result of positive contributions for the age groups under 15 and 45 or more years old, which more than offset negative contributions for the age groups from 15 through 44 years old. Similar positive and negative contributions occurred for White males and White females (data not shown). For Black males (data not shown), life expectancy declined in the period from 1984 through 1989 by −0.730 year, reflecting negative contributions for ages under 5, 10–54, 75–79, and 85 years and over, offsetting positive contributions for other age groups. For Black females (data not shown), life expectancy declined −0.178 year, reflecting negative contributions occurring at ages under 5, 10–45, 65–69, 75–79, and 85 years and over, also offsetting positive contributions in the other age groups.

By Cause of Death

For the total population, the cause of death with the largest positive contribution to the change in life expectancy in the period from 1984 through 1989 (Table 2, "By Cause") was heart disease, followed by cerebrovascular disease (stroke). These positive contributions were partially offset by negative contributions for other causes of death, the largest of which were for HIV infection, diabetes mellitus (diabetes), and pneumonia and influenza (pneumonia).

For White males (data not shown), heart disease contributed the most, followed by stroke and accidents and adverse effects (accidents). For White females the pattern was similar to that of White males.

Although life expectancy increased from 1984 through 1989 for both White males and females, it declined for Black males and females. For Black males (data not shown), the greatest negative contributor was HIV infection, followed by homicide and legal intervention (homicide). For Black females, the major causes of the decrease in life expectancy were HIV infection, diabetes, and pneumonia.

By Age and Cause of Death

A more comprehensive way to examine contributions to the change in life expectancy is to combine causes of death and age groups. Thus, for some causes of death (e.g., cancer) there were positive contributions for some age groups and negative contributions for other age groups.

For the total population, over half (59.4%) of the positive contribution to the change in life expectancy from 1984 through 1989 was due to changes in mortality from heart disease for ages birth through 1 year and 35 years and over (Table 2, "By Age and Cause"). Also making a large contribution were stroke for ages 40 years and over and cancer for ages 30–59 years. These three causes of death accounted for over three fourths of the positive contribution to the change in life expectancy during this period. Negative contributions for the total population were largely attributable to HIV infection, which accounted for one third of the negative contribution to the change in life expectancy in the period from 1984 through 1989. For this cause, the negative contribution was concentrated in persons aged 20–59 years. Cancer for ages 65 years and over accounted for 9.3% of the negative contribution. Pneumonia for ages 75 years and over accounted for 8.6% of

TABLE 2—Rank Ordering of the Contribution to the Change in Life Expectancy at Birth: Total Population, United States, 1984 through 1989

By Age			By Cause			By Age and Cause			
Age, y	Change in Life Expectancy, y	%	Cause of Death	Change in Life Expectancy, y	%	Cause of Death	Major Contributing Age Groups,[a] y	Change in Life Expectancy,[b] y	%
				Total contribution					
All ages	0.513	. . .	All causes	0.513	. . .	All causes	All	0.513	. . .
				Positive contribution					
Total	0.665	100.0	Total	0.942	100.0	Total	. . .	1.057	100.0
70–74	0.097	14.6	Heart disease	0.627	66.6	Heart disease	0–1,35+	0.6 7	59.4
65–69	0.092	13.8	Stroke	0.131	13.9	Stroke	40+	0. ..	12.4
60–64	0.084	12.6	Perinatal conditions	0.037	3.9	Cancer	30–59	0.067	6.3
55–59	0.076	11.4	Atherosclerosis	0.035	3.7	Accidents	1–4, 10–14, 20–29, 45–49	0.039	3.7
<1	0.071	10.7	Accidents	0.033	3.5				
50–54	0.067	10.1	Congenital anomalies	0.025	2.6	Perinatal conditions	0–1	0.037	3.5
75–79	0.066	9.9	Liver disease	0.025	2.6	Atherosclerosis	75+	0.035	3.3
80–84	0.040	6.0	Cancer	0.016	1.7	Congenital anomalies	0–1	0.027	2.5
45–49	0.033	4.9	Nephritis	0.009	0.9	Liver disease	45–59, 65–69	0.026	2.5
85+	0.023	3.4	Suicide	0.005	0.6	Suicide	. . .	0.013	1.3
1–4	0.007	1.1				Nephritis	. . .	0.009	0.9
5–9	0.006	0.8				Residual	. . .	0.045	4.3
10–14	0.003	0.5							
				Negative contribution					
Total	−0.152	100.0	Total	−0.429	100.0	Total	. . .	−0.543	100.0
35–39	−0.051	33.7	HIV infection[c]	−0.192	44.8	HIV infection[c]	20–59	−0.192	35.4
30–34	−0.046	30.5	Diabetes	−0.045	10.6	Cancer	65+	−0.051	9.3
25–29	−0.028	18.2	Pneumonia	−0.044	10.2	Pneumonia	75+	−0.047	8.6
15–19	−0.019	12.2	COPD	−0.040	9.2	Diabetes	45+	−0.045	8.3
40–44	−0.005	3.4	Homicide	−0.034	8.0	COPD	60–64, 70+	−0.040	7.4
20–24	−0.003	2.0	Septicemia	−0.012	2.8	Homicide	15–34	−0.040	7.3
			Hypertension	−0.001	0.2	Septicemia	85+	−0.013	2.5
			Residual	−0.061	14.3	Suicide	15–19	−0.008	1.4
						Accidents	35–39	−0.006	1.1
						Hypertension	. . .	−0.002	0.4
						Residual	. . .	−0.099	18.3

Note. Data are based on recomputed life expectancies at birth rounded to the nearest 0.001 year. COPD = chronic obstructive pulmonary disease.
[a]Includes age groups that contributed at least 1 day to the change in life expectancy.
[b]May include positive (or negative) changes for age groups not shown.
[c]Number of resident deaths due to human immunodeficiency virus infection in 1984 was estimated by using the number of death certificates with mention of ICD-9 code 279.1.

the negative contribution. Diabetes also showed a negative contribution for ages 45 years and over.

Causes of death may show sizable positive contributions for some age groups and sizable negative contributions for other age groups. This is particularly the case for cancer. Decreasing mortality from cancer for persons aged 30–59 years made a 6.3% positive contribution to the change in life expectancy from 1984 through 1989, whereas for persons aged 65 years and over, increasing mortality from this cause contributed 9.3% to the negative change in life expectancy. However, for most causes of death, the contributions of the age groups tend to be generally positive or negative.

White males showed a pattern similar to that of the total population (Table 3), with a total positive contribution of 59.1% from the age-cause combinations of heart disease for ages under 1 year and 35 years and over; stroke for ages 45 years and over (7.8%); and accidents for ages 1–4, 10–34, and 45–54 years (7.0%). Negative contributions to life expectancy were dominated by HIV infection for ages 20–64 years, representing 55.6% of all negative contributions.

For White females (Table 3), causes of death making large positive contributions were heart disease for ages under 1 year and 40 years and over, stroke for ages 40 years and over, and cancer for ages 30–59 years. Dominating the negative contributions for White females (unlike White males) were chronic obstructive pulmonary diseases and allied conditions for ages 55 years and over, cancer for ages 60 years and over, and pneumonia for ages 65–69 years and 75 years and over.

For Black males (Table 4), heart disease for ages under 1 year and 25 years and over had a major positive impact, as

TABLE 3—Rank Ordering of the Contribution of Leading Causes of Death to the Change in Life Expectancy at Birth for White Males and Females: United States, 1984 through 1989

White Males				White Females			
Cause of Death	Major Contributing Age Groups,[a] y	Change in Life Expectancy,[b] y	%	Cause of Death	Major Contributing Age Groups,[a] y	Change in Life Expectancy,[b] y	%
		Total contribution					
All causes	All	0.831	...	All causes	All	0.499	...
		Positive contribution					
Total	...	1.306	100.0	Total	...	0.981	100.0
Heart disease	0–1, 35+	0.772	59.1	Heart disease	0–1, 40+	0.515	52.5
Stroke	45+	0.101	7.8	Stroke	40+	0.159	16.2
Accidents	1–4, 10–34, 45–54	0.092	7.0	Cancer	30–59	0.060	6.1
Cancer	15–19, 30–34, 40–59, 65–69	0.077	5.9	Atherosclerosis	75+	0.048	4.9
				Perinatal conditions	0–1	0.047	4.8
Perinatal conditions	0–1	0.059	4.6	Congenital anomalies	0–1	0.030	3.1
Congenital anomalies	0–4	0.034	2.6	Liver disease	40–59	0.023	2.4
Liver disease	45–69	0.027	2.1	Suicide	20–24, 40–49	0.023	2.3
Atherosclerosis	75+	0.023	1.8	Accidents	20–24	0.016	1.6
COPD	65–74	0.020	1.5	Nephritis	...	0.010	1.0
Suicide	55–59	0.013	1.0	Residual	...	0.050	5.1
Residual	...	0.087	6.7				
		Negative contribution					
Total	...	−0.476	100.0	Total	...	−0.482	100.0
HIV infection[c]	20–64	−0.264	55.6	COPD	55+	−0.098	20.3
Diabetes	50–74	−0.035	7.3	Cancer	60+	−0.085	17.6
Cancer	60–64, 80+	−0.028	5.9	Pneumonia	65–69, 75+	−0.067	13.9
Pneumonia	80+	−0.022	4.6	Diabetes	50+	−0.045	9.4
Suicide	15–19	−0.014	2.9	HIV infection[c]	25–39	−0.023	4.8
Homicide	15–24	−0.011	2.3	Septicemia	80+	−0.018	3.8
COPD	85+	−0.008	1.7	Accidents	15–19	−0.017	3.5
Septicemia	...	−0.008	1.7	Homicide	...	−0.004	0.8
Liver disease	...	−0.003	0.7	Congenital anomalies	...	−0.003	0.7
Accidents	...	−0.003	0.6	Hypertension	...	−0.003	0.6
Residual	...	−0.080	16.7	Residual	...	−0.119	24.6

Note. Data are based on recomputed life expectancies at birth rounded to the nearest 0.001 year. COPD = chronic obstructive pulmonary disease.
[a]Includes age groups that contributed at least 1 day to the change in life expectancy.
[b]May include positive (or negative) changes for age groups not shown.
[c]Number of resident deaths due to human immunodeficiency virus infection in 1984 was estimated by using the number of death certificates with mention of ICD-9 code 279.1.

did cancer for ages 30–49 and 55–64 years and stroke for ages 30–34 and 40 years and over. A striking effect was that of homicide at ages 45–64 years, which contributed positively to the change in life expectancy. Negative effects for Black males were dominated by HIV infection for ages under 5 and 20–69 years; homicide for ages 1–4 and 10–34 years; and accidents for ages under 1 year and 15–24, 35–54, and 60–64 years.

For Black females (Table 4), three fourths of the positive contribution to the change in life expectancy from 1984 through 1989 was due to the three leading causes of death: heart disease for ages under 5 and 35–84 years, stroke for ages 35 years and over, and cancer for ages 25–29 and 40–59 years. Negative effects were led by HIV infection for ages under 5 and 20–59 years, cancer for ages 65–79 and 85 years and over, and diabetes for ages 45–69 and 75 years and over.

Change in the Gap in Life Expectancy between Whites and Blacks

Historically, Black life expectancy has, on the average, increased more rapidly than White life expectancy, resulting in a narrowing over time in the gap in average length of life between the two groups. But in the period from 1984 through 1989 the gap or difference in life expectancy between White and Black males increased by 1.7 years from 6.2 years to 7.9 years (Table 1). Contributing most to the growing gap in life expectancy between males of the two race groups were trends in mortality for heart disease for males 40 years and over (Figure 1). Although the change in life expectancy from this cause increased for both White and Black males, it increased less rapidly for Black males, resulting in a negative contribution of 23.2% of the widening gap. The next two most important causes contributing to the widening gap were HIV infection (for ages under 5 and 20–64 years) and homicide (for ages 1–4 and 10–39 years), for which decreases in the change in life expectancy were greater for Black males than for White males, resulting again in negative contributions to the gap in life expectancy.

Some causes of death and age groups kept the gap between Black and White males from widening more than it did

TABLE 4—Rank Ordering of the Contribution of Leading Causes of Death to the Change in Life Expectancy at Birth for Black Males and Females: United States, 1984 through 1989

Black Males				Black Females			
Cause of Death	Major Contributing Age Groups,[a] y	Change in Life Expectancy,[b] y	%	Cause of Death	Major Contributing Age Groups,[a] y	Change in Life Expectancy,[b] y	%
		Total contribution					
All causes	All	−0.730	. . .	All causes	All	−0.178	. . .
		Positive contribution					
Total	. . .	0.825	100.0	Total	. . .	0.755	100.0
Heart disease	0–1, 25+	0.350	42.4	Heart disease	0–4, 35–84	0.299	39.6
Cancer	30–49, 55–64	0.135	16.3	Stroke	35+	0.150	19.9
Stroke	30–34, 40+	0.111	13.4	Cancer	25–29, 40–59	0.112	14.8
Liver disease	25–39, 45–59	0.038	4.6	Liver disease	25–44, 50–59	0.044	5.8
Homicide	45–64	0.034	4.1	Atherosclerosis	80+	0.024	3.2
Accidents	1–14, 30–34, 55–59, 70–74	0.030	3.6	Congenital anomalies	0–1, 5–9	0.022	3.0
Atherosclerosis	65+	0.019	2.3	Nephritis	0–1, 55–59, 70–74	0.018	2.4
COPD	45–49, 60–64	0.011	1.4	Accidents	1–4, 50–54	0.015	2.0
Nephritis	0–1, 60–64	0.011	1.3	Septicemia	55–59	0.008	1.0
Pneumonia	0–1	0.010	1.2	Homicide	5–9	0.008	1.0
Residual	. . .	0.078	9.4	Residual	. . .	0.055	7.3
		Negative contribution					
Total	. . .	−1.556	100.0	Total	. . .	−0.934	100.0
HIV infection[c]	0–4, 20–69	−0.664	42.7	HIV infection[c]	0–4, 20–59	−0.188	20.2
Homicide	1–4, 10–34	−0.329	21.1	Cancer	65–79, 85+	−0.113	12.1
Accidents	0–1, 15–24, 35–54, 60–64	−0.078	5.0	Diabetes	45–69, 75+	−0.097	10.4
Diabetes	25–29, 45–79, 85+	−0.069	4.4	Pneumonia	30–39, 45–49, 70+	−0.083	8.8
Cancer	20–24, 65+	−0.065	4.2	COPD	50+	−0.067	7.2
Pneumonia	35–44, 65–74, 80+	−0.049	3.2	Accidents	0–1, 5–9, 25–44, 65–69, 85+	−0.060	6.4
Suicide	15–24, 30–34, 40–44	−0.040	2.6				
COPD	30–34, 40–44, 65–69, 75–79, 85	−0.039	2.5	Homicide	0–1, 15–39	−0.051	5.5
				Heart disease	85+	−0.030	3.2
Nephritis	30–44, 50–59	−0.025	1.6	Septicemia	75–79, 85+	−0.029	3.1
Septicemia	35–39	−0.018	1.1	Perinatal conditions	0–1	−0.017	1.8
Residual	. . .	−0.181	11.7	Residual	. . .	−0.199	21.3

Note. Data are based on recomputed life expectancies at birth rounded to the nearest 0.001 year. COPD = chronic obstructive pulmonary disease.
[a]Includes age groups that contributed at least 1 day to the change in life expectancy.
[b]May include positive (or negative) changes for age groups not shown.
[c]Number of resident deaths due to human immunodeficiency virus infection in 1984 was estimated using the number of death certificates with mention of ICD-9 code 279.1.

(Figure 1). Among these were trends in mortality for cancer for males aged 35–49 and 55–64 years, which made a positive contribution of 32.2% to narrowing the gap. Also contributing to the narrowing were stroke for ages 55–59 and 65–74 years (11.6%) and homicide for ages 45–64 years (9.5%).

From 1984 through 1989 the gap in life expectancy between White and Black females also increased, by 0.7 year, from 5.0 years to 5.7 years (Table 1). Contributing most to the growing gap in life expectancy between females of the two race groups were trends in mortality for heart disease for ages 60 years and over (Figure 2). Although the change in life expectancy from this cause increased for females of both races, it increased less rapidly for Black females, resulting in a negative contribution of 27.5% to the widening gap. For the next most important cause contributing to the widening gap—namely, HIV infection (for ages under 5 years and 20–64 years)—decreases in the change in life expectancy were greater for Black females than for White females, resulting in a negative contribution to the widening gap in life expectancy. These two causes in the age groups mentioned contributed 43.2% to the widening gap in life expectancy between White and Black females.

Countervailing effects to the widening of the gap were mortality trends for cancer for females aged 25–29, 40–44, 50–64, 70–74, and 80–84 years, which contributed to the narrowing of the gap. Also contributing to the narrowing were chronic obstructive pulmonary disease for ages 60–64 and 70 years and over and heart disease for ages 35–44 and 50–59 years.

Discussion

A major change occurred in US patterns of mortality from 1984 through 1989: whereas life expectancy for the total population continued the general increase that has occurred since the beginning of the century, for the very first time there was evidence of strong and persistent divergence in the mortality patterns

between the two major race groups, the White and the Black populations. Life expectancy increased for White men and women but decreased for Black men and women. Never before has this pattern occurred in the United States over a sustained period. Previously, mortality patterns for the two major race groups have been upward, with the Black death rates improving more rapidly than the White rates. This trend has resulted, historically, in a convergence in mortality between the two groups. But during the period from 1984 through 1989, this pattern was reversed.

The results show that from 1984 through 1989 the trends and patterns in mortality between the White and Black populations were sufficiently different to result in differences in the change in life expectancy. For the White population, improvements in heart disease, stroke, and accident mortality were sufficient to offset the negative trends of other causes. However, for the Black population, the negative impact of HIV infection, homicide, diabetes, and pneumonia overwhelmed improvements in mortality from other causes.

It should be noted, and may come as a surprise, that the greatest contributions to the widening gap in life expectancy between the two major race groups were made by a cause of death for which the trends are favorable: heart disease. For both race groups, mortality from this cause is declining, but it is declining so much more rapidly for the White than for the Black population that this is causing the gap to widen.

Obviously, changes in mortality patterns are highly complex, particularly when examined separately for males and females of each of the major race groups and then by age and cause of death. A major contribution to articulating and understanding these changes is the life table partitioning method developed by Eduardo E. Arriaga. This method provides a clear way to define in demographic and medical terms the major mortality problems in the United States and to suggest where prevention efforts might have their greatest impact.

This analysis has focused on changes in mortality by race, a variable that is widely used in health statistics to classify the population. Many studies suggest that race may be a proxy, in part or substantially, for social and economic factors such as educational attainment, income, occupation, and life-style.[7,8] As such, the results described in this article may be indicative of changes occurring in the social fabric of society regardless of race, rather than being specific to one racial group or another. Our results are consistent with those of Pappas et al.,[9] who documented growing gaps in mortality by socioeconomic status. Such findings are consistent with and may be associated with the increasing gap in life expectancy between the White and Black populations in the period from 1984 through 1989.

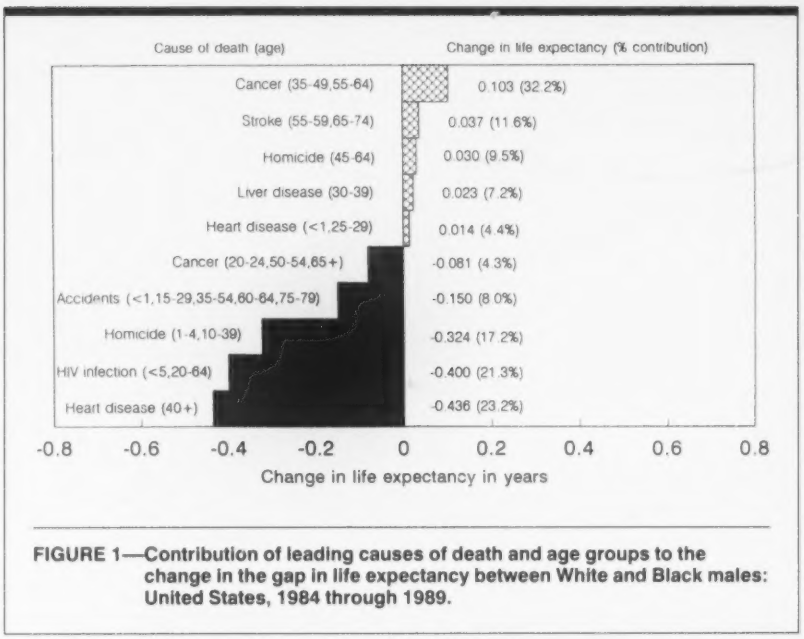
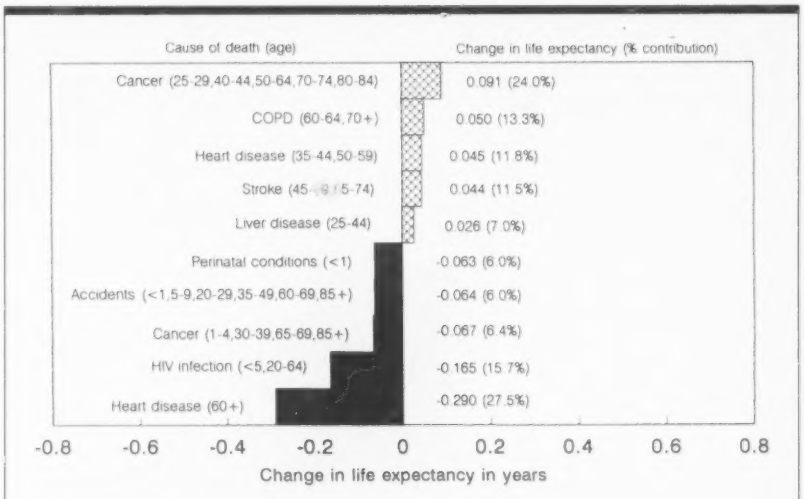


FIGURE 1—Contribution of leading causes of death and age groups to the change in the gap in life expectancy between White and Black males: United States, 1984 through 1989.



Note. COPD = chronic obstructive pulmonary disease.

FIGURE 2—Contribution of leading causes of death and age groups to the change in the gap in life expectancy between White and Black females: United States, 1984 through 1989.

Some important health problems are shared by all Americans, but for some problems the burden falls more heavily on one racial group than another, on one gender more than the other, and on some age groups more than others. It is clear that not all groups have shared equally in the improvements in health reflected in US mortality statistics, particularly from 1984 through 1989, when the gap in health between our White and Black populations

widened in an unprecedented manner. This analysis identifies the specific ways in which this widening occurred. □

References

1. National Center for Health Statistics. Advance report of final mortality statistics, 1989. *Monthly Vital Stat Rep.*; 1992; 40(8)(suppl 2).
2. Arriaga EE. Measuring and explaining the change in life expectancies. *Demography.* 1984;21:83–96.
3. Arriaga EE. Changing trends in mortality decline during the last decades. In: Ruzicka L, Wunsch G, Kane P, eds. *Differential Mortality: Methodological Issues and Biosocial Factors.* Oxford, England: Clarendon Press; 1989:105–129.
4. Pollard JH. On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography.* 1988;25:265–276.
5. *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death,* Ninth Revision, 1975. Geneva, Switzerland: World Health Organization; 1977.
6. *Vital Statistics of the United States, 1988.* Vol. 2, Mortality, Part A. Hyattsville, Md: National Center for Health Statistics; 1991. DHHS publication PHS 91–1101.
7. Otten MW Jr, Teutsch SM, Williamson DF, Marks JS. The effects of known risk factors on the excess mortality of black adults in the United States. *JAMA.* 1990;263:845–850.
8. Kleinman JC. The slowdown in the infant mortality decline. *Paediatr Perinat Epidemiol.* 1990;4:373–381.
9. Pappas G, Queen S, Hadden W, Fisher G. The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *N Engl J Med.* 1993;329:103–109.

APPENDIX—Causes of Death Used in Study (ICD-9 Codes)

1. Diseases of heart (390–398, 402, 404–429) (heart disease)
2. Malignant neoplasms, including neoplasms of lymphatic and hematopoietic tissues (140–208) (cancer)
3. Cerebrovascular diseases (430–438) (stroke)
4. Accidents and adverse effects (E800–E949) (accidents)
5. Chronic obstructive pulmonary diseases (COPD) and allied conditions (490–496)
6. Pneumonia and influenza (480–487) (pneumonia)
7. Diabetes mellitus (250) (diabetes)
8. Suicide (E950–E959)
9. Chronic liver disease and cirrhosis (571) (liver disease)
10. Homicide and legal intervention (E960–E978) (homicide)
11. Human immunodeficiency virus infection (*042–*044) (HIV infection)[a]
12. Nephritis, nephrotic syndrome, and nephrosis (580–589) (nephritis)
13. Atherosclerosis (440)
14. Septicemia (038)
15. Certain conditions originating in the perinatal period (760–779) (perinatal conditions)
16. Hypertension with or without renal disease (401, 403) (hypertension)
17. Congenital anomalies (740–759)
18. All other causes (residual)

[a]Not part of ICD-9; added by the United States in 1987.

The Effects of Occupational Class Transitions on Hypertension: Racial Disparities among Working-Age Men

ABSTRACT

Objectives. This study was undertaken to examine how the interaction between occupational class transitions and race affects the incidence of hypertension.

Methods. A cohort of 1982 men (183 Black), ages 25 to 55, received a baseline medical exam between 1971 and 1975 and a follow-up between 1982 and 1984. Logistic regressions were estimated for hypertension at follow-up controlling for hypertension at baseline, other risk factors associated with blood pressure, and interaction terms identifying specific occupational class transitions among Blacks and Whites. The occupational class matrix was based largely on scores of US Census Bureau occupations from the *Dictionary of Occupational Titles.*

Results. Relative to Whites who remained in professional and technical jobs between baseline and follow-up, Blacks and Whites who remained in lower occupational classes or made specific transitions—notably into the lowest class—had significantly higher incidence rates of hypertension. These differences were greater among Blacks, who are also more concentrated in and less likely to move upward from the lower end of the occupational class matrix.

Conclusions. Widening racial disparities in high blood pressure over the period of study may be partly attributable to characteristics associated with occupational class position and dynamics. (*Am J Public Health.* 1994;84:945–950)

Norman J. Waitzman, PhD, and Ken R. Smith, PhD

Introduction

Hypertension is a major risk factor for heart disease and stroke.[1] Its higher prevalence among Blacks relative to Whites in the United States is well documented[2–4] although the mechanisms underlying this racial disparity are not fully understood. The distinct income and social class gradients exhibited by blood pressure[3–6] suggest that characteristics associated with the disproportionate concentration of Blacks in poverty and low-paying, insecure jobs are partly responsible.

Several epidemiological studies support the view that psychosocial stress among lower socioeconomic strata contributes to differences in blood pressure,[4,6–9] and a review of the literature suggests that employment security and other work conditions play a pivotal role.[10] The presence of low decision latitude on the job in the face of high demands, a pattern most characteristic of jobs with low social status, has been shown to be directly associated with the risk of cardiovascular disease[11] and elevated blood pressure.[12–14] A separate group of studies on blood pressure by race[7–9] has focused on internalized demands arising from a belief system that has been labeled "active coping" or "John Henryism": the perception that environment can be mastered through individual determination. These studies have generally supported an interactive relationship between John Henryism and low socioeconomic status leading to elevated blood pressure among Blacks[7–9] but not among Whites.[9]

Most of the above research relating psychosocial stress to blood pressure has drawn inferences from cross-sectional data[7–9,12] and has focused on particular subgroups of the population, such as residents of rural, southern communities[7–9]; workers within a very limited number of occupations[13] or occupational strata[7–9]; or a cohort of relatively young workers.[14] The few longitudinal studies that have investigated the association between occupational strain and blood pressure[13–14] are limited with respect to sample size and scope, as noted above, and have not addressed racial differences. Furthermore, none has investigated the experience of social mobility as a potential factor. Some research has suggested that the prospects of upward occupational mobility are greater for Whites than for Blacks.[15]

We performed a longitudinal analysis of occupational class transitions and blood pressure changes among a nationally representative sample of Black and White men in the United States who were followed for 7 to 13 years after receiving a baseline medical examination. Our measures of occupational class were based on characteristics that have distinct similarities to the demand/latitude characteristics in earlier research.[11–14] While our data do not permit us to include explicit measures of active coping in our model, some of the significant correlates of John Henryism in earlier studies, such as education and occupational status, figure into our classification scheme.

Methods

Study Sample

The sample consisted of a cohort of Black and White men from the first

Norman J. Waitzman is with the Department of Economics and Ken R. Smith is with the Department of Family and Consumer Studies, University of Utah, Salt Lake City.

Requests for reprints should be sent to Norman J. Waitzman, PhD, Department of Economics, BUC 308, University of Utah, Salt Lake City, UT 84112.

This paper was accepted February 15, 1994.

National Health and Nutrition Examination Survey (baseline) and the National Health Examination Follow-Up Survey (follow-up) who were aged 25 to 55 at baseline; were neither retired nor unable to work owing to sickness at baseline or follow-up; and had complete blood pressure, occupational, and other covariate data. Men aged 56 or older were excluded from the sample because of their low rates of occupational reporting at follow-up largely owing to retirement. The baseline survey was conducted between 1971 and 1975, and the follow-up was conducted between 1982 and 1984. Of 2666 males at baseline who met the above criteria, 149 died and 218 were otherwise lost to follow-up. Missing data for occupational and other variables were responsible for an additional loss of 317 respondents. Of the remaining 1982 respondents who met all the sample selection criteria, 183 were Black.

The potential for sample selection bias is of particular concern in longitudinal panel studies because those lost to follow-up may differ in some nonrandom way from those who remain in the sample.[16] Sample selection bias in the regressions was formally tested and corrected for, based on the Heckman technique.[16,17] The test suggested the absence of such bias, so results are reported without the sample selection bias correction.

Specification of Variables

The measurement protocol for blood pressure was similar but not identical across surveys. A single measurement according to then-standard guidelines[3] was taken at baseline while three successive measures during a single exam were taken at follow-up. Summary data and regression results reported below use the first reading at follow-up to preserve consistency in measurement across the two surveys. Results from regressions using averages of the second and third readings at follow-up in accordance with current guidelines,[2] however, were nearly identical to those reported here.

Both dichotomous measures of hypertension used in this analysis are based on definitions of the lower bounds for mild hypertension (diastolic ≥ 90 mm Hg) or borderline systolic hypertension (systolic ≥ 140 mm Hg) established by the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure.[2] The second measure also incorporates information on usage of blood pressure medication. Respondents were asked whether they were taking prescribed blood pressure medication at follow-up. Information on such medication at baseline was gathered for a nationally representative subsample comprising two thirds of the baseline respondents, however, and therefore was not used in the analysis to ensure as large a sample as possible. In analyses based on this subsample, such an adjustment made no significant difference in the results.

Respondents were overweight if their body mass index (weight in kilograms/height in meters squared) at baseline equaled or exceeded that of the 85th percentile of the body mass distribution among the adult male population between 1976 and 1980.[18] The variable for daily alcohol consumption was based on self-reports of the amount and frequency of consumption at baseline and was calculated as number of drinks per day $\times$ reported frequency of drinking $\times$ 0.5, with 0.5 being an estimate of the amount of ethanol in a shot of liquor, a 5-oz glass of wine, or a 12-oz glass of beer.[19]

Occupational Classification

Health and labor market studies that use a scheme for occupational status have often relied on census one-digit occupational categories. However, these categories tend to be heterogeneous, give insufficient weight to job content and its changes over time, and lack a consistent base in labor market or social theory.[20,21] Drawing on labor market segmentation theory to overcome these weaknesses,[22,23] this analysis used occupational ratings at the three-digit level for census occupations and incorporated differences in jobs arising from industry characteristics.

According to segmentation theorists, the labor market is roughly split between a primary market in which jobs are relatively secure, pay well, and offer opportunities for training and advancement, and a secondary market characterized by low-paying, dead-end jobs. For this study, a four-part segmentation scheme, which partitions the primary labor market into three segments, was adopted.[23,24] The topmost segment (class 1) comprises technical and professional jobs that require several years of general educational preparation and involve independent decision-making and/or supervisory responsibilities. Jobs in class 2 are craft jobs that require several years of vocational training and also afford latitude in decision making. Occupations in class 3 are more often machine paced and subject to strict timetables, rules, and supervision than are jobs in the upper two classes, but they still offer job security and opportunities for advancement. Such security and opportunities in secondary labor market occupations (class 4) are minimal even though skill requirements may be similar to those for entry-level jobs in class 3.

The four-part segmentation scheme was operationalized on the basis of both scores on educational and training requirements for occupations and ratings on the extent to which jobs afford decision-making latitude or require working at a repetitious or set pace, as per the *Dictionary of Occupational Titles*.[25] Workers could be reassigned between the bottom two classes based on census industrial characteristics. A description of the method is available from another source.*

Occupational Variables

A respondent could be in any of the four occupational classes at baseline and at follow-up, suggesting a 4 $\times$ 4 class transition matrix. Cells along the diagonal of such a matrix identify those who were in the same class in both surveys, whereas off-diagonal cells identify "movers." Because the number of Blacks in the sample in several off-diagonal cells was small, it was not possible to construct variables specifying all 16 occupational transitions by race. Accordingly, certain cells were collapsed for use in subsequent regressions. In particular, the direction of an occupational move (up or down) and the destination (class at follow-up) are the two dimensions used to identify occupational transitions. Destination classes are important because they are more proximal in time to recorded levels of blood pressure at follow-up than are classes of origin. Downward moves may generally be considered more detrimental to health than upward moves.

Dummy variables identify specific transitions that preserve direction and destination for both Whites and Blacks. The variable "Moved down to class 4," for example, identifies downward moves from

*See National Auxiliary Publication Service (NAPS) document 05095 for 7 pages of Appendix—Procedure for Allocating Workers by Occupational Segment. Please order directly from NAPS c/o Microfiche Publications, PO Box 3513, Grand Central Station, New York, NY 10163-3513. Enclose with your order (a) $7.75 for paper copy or (b) $4.00 for microfiche (US only funds from a US bank). For orders from outside the United States and Canada, add postage of $4.50 for paper copy or $1.50 for microfiche. There is a $15.00 invoicing charge for all orders filled before payment.

class 1, 2, or 3. The "Stayed in class 1, 2, 3, or 4" dummies identify the four groups that did not change occupational class.

The model from which regression results are reported below incorporates interaction terms between race and occupational class transitions. Goodness-of-fit comparisons between logistic regressions based on changes in the likelihood ratio indicate that the interaction terms detected in the final model fit the data better than simpler models. These results suggest that, in addition to theoretical considerations, this is the "best" model on statistical grounds.[26,27]

Model Specification

Logistic regressions were run on two dichotomous measures of hypertension at follow-up. One measure is an indicator of uncontrolled hypertension; the other counts as hypertensive an additional 150 respondents on antihypertensive medication whose blood pressure was controlled at follow-up. The two measures were incorporated to test the extent to which hypertension control through medication was related to occupational class transition.

Use of blood pressure medication at follow-up was incorporated as a separate covariate in the regression on uncontrolled hypertension. In this regression, such medication was expected to have a positive relation to changes in blood pressure, as those who took it would have done so because their blood pressure was increasing. However, among hypertensives at baseline, those taking medication at follow-up were expected to have lower blood pressure than their nonmedicated counterparts. Blood pressure medication was therefore also interacted with the baseline measure of hypertension in this regression, and these expectations were borne out in the results.

The association between occupational class and high blood pressure was expected to follow a general inverse relationship.[3-6] The defining characteristics of the occupational segmentation scheme focused the analysis on underlying mechanisms for this relationship, such as demands imposed by machine pacing and the level of decision-making authority, as suggested by the literature on occupational characteristics and blood pressure.[10,12-14] Staying in or moving into lower rather than higher occupational classes was therefore expected to be associated with the largest increases in blood pressure.

TABLE 1—Sample Distribution by Occupational Class[a] at Baseline and Follow-Up and by Occupational Class Transition, by Race[b]

	Baseline		Follow-Up		Transition	
	White (n = 1799)	Black (n = 183)	White (n = 1799)	Black (n = 183)	White (n = 1799)	Black (n = 183)
Class						
1	764 (43.8)	33 (15.0)	764 (43.9)	31 (15.3)	...	...
2	275 (14.7)	27 (17.9)	274 (15.1)	26 (12.7)	...	...
3	402 (23.8)	39 (20.5)	388 (21.7)	48 (29.8)	...	...
4	358 (17.8)	84 (46.5)	373 (19.3)	78 (42.3)	...	...
Stayed in class						
1	...	...	...	...	575 (33.3)	22 (10.4)
2	...	...	...	...	163 (8.9)	14 (8.8)
3	...	...	...	...	212 (12.6)	23 (16.5)
4	...	...	...	...	209 (10.4)	57 (32.9)
Moved down to class						
2	...	...	...	...	48 (2.8)	1 (0.2)
3	...	...	...	...	120 (6.5)	10 (5.3)
4	...	...	...	...	164 (8.9)	21 (9.4)
Moved up to class						
1	...	...	...	...	189 (10.6)	9 (4.9)
2	...	...	...	...	63 (3.5)	11 (3.8)
3	...	...	...	...	56 (2.6)	15 (8.0)

[a]In general, jobs in class 1 rank high in educational requirements and/or decison-making authority; jobs in class 2 rank high with respect to vocational training; jobs in class 3 rank low with respect to vocational training and/or high in repetitive operations; jobs in class 4 rank very low with respect to training requirements.
[b]Weighted column percentages in parentheses may not add to 100 because of rounding.

The baseline measure of hypertension was incorporated as a covariate in each regression. Other covariates, such as overweight, alcohol intake, baseline age, and aging (the time between baseline and follow-up exams), were incorporated because of their demonstrated association with elevated blood pressure.[2] Population density has an inverse association with measures of blood pressure,[3] perhaps because of better access to medical care or medical information in urbanized areas. Thus, an indicator variable was incorporated for residence in the center city of a standard metropolitan statistical area as a proxy for access to general medical resources.

Results

The distribution of the study sample by occupational class and race at baseline and follow-up reported in the first four columns of Table 1 is similar to that reported in other segmentation studies,[15,24] with Blacks more than Whites being concentrated in the lower reaches of the occupational hierarchy (classes 3 and 4) at both baseline and follow-up. Whites are represented more in professional and technical jobs (class 1).

The distribution of the sample by transition is reported in the final two columns of the table. About two thirds of both Black and White workers remained in the same occupational class between baseline and follow-up while one third moved. Among stayers, however, about half of Blacks stayed in the bottom class while nearly half of Whites remained in the top class. Among those outside the top class at baseline, a greater percentage of Whites than of Blacks moved up, particularly into the top class, whereas among those outside the bottom class, a greater percentage of Blacks than of Whites moved down, particularly into the bottom class. This mobility pattern is consistent with that found in other research.[15]

Age-adjusted means for blood pressure at baseline and follow-up by race and occupational transition are presented in Table 2. The decline in these means

TABLE 2—Mean (Age-Adjusted) Baseline and Follow-Up Blood Pressure, by Occupational Class[a] Transition and Race

	Systolic Blood Pressure (mm Hg)				Diastolic Blood Pressure (mm Hg)			
	White		Black		White		Black	
	Base-line	Follow-Up	Base-line	Follow-Up	Base-line	Follow-Up	Base-line	Follow-Up
Stayed in class								
1	125.4	124.5	128.4	127.9	82.5	78.9	89.3	78.7
2	128.9	127.2	127.4	144.9	84.4	79.3	83.2	92.3
3	130.2	129.0	130.6	134.3	86.2	79.9	88.4	80.7
4	128.6	128.8	131.0	134.8	84.2	79.7	87.0	87.4
Moved down to class								
2	131.6	125.1	. . .	. . .[b]	87.3	81.8	. . .	. . .[b]
3	127.2	127.1	136.0	130.1	84.1	79.5	82.0	75.5
4	126.3	127.3	133.7	145.0	83.7	79.5	89.2	90.2
Moved up to class								
1	129.4	127.4	136.3	124.6	86.0	80.7	96.1	81.2
2	125.7	127.9	133.6	131.3	84.3	80.9	85.2	79.7
3	127.0	125.2	137.9	135.3	84.6	78.7	93.7	93.9

[a]In general, jobs in class 1 rank high in educational requirements and/or decison-making authority; jobs in class 2 rank high with respect to vocational training; jobs in class 3 rank low with respect to vocational training and/or high in repetitive operations; jobs in class 4 rank very low with respect to training requirements.
[b]Only one case in this category.

between baseline and follow-up for several of the transitions for both races reflects, to a certain extent, the greater use of blood pressure medication at follow-up (about 16%) than at baseline (about 3% based on a two-thirds random sample).

Odds ratios and 95% confidence intervals from the logistic regressions on hypertension for the variables identifying class and race interactions are reported in Table 3. The reference category is "stayed in class 1" among Whites. Results from equation 1 indicate that Whites and Blacks who stayed in class 3 or 4 had a greater risk of developing uncontrolled hypertension than did Whites remaining in class 1, although the odds ratios for Blacks were more than twice that of their White counterparts. An even more profound difference surfaced for Blacks remaining in class 2, although the generally wider confidence intervals on results for Blacks, reflecting smaller numbers, are particularly pronounced for this transition. Blacks remaining in class 1 did not exhibit a significant difference from their White counterparts.

Among those experiencing occupational class transitions, downward movement into the bottommost class was associated with significantly higher odds of uncontrolled hypertension among Blacks. The relative risk was also significantly higher for certain upward transitions, most notably among Blacks moving from class 4 to class 3.

Results (not shown) on a measure of more severe uncontrolled hypertension (follow-up diastolic ≥ 105 mm Hg or follow-up systolic ≥ 160 mm Hg) indicated that the odds ratios for the above patterns were generally enhanced among both Blacks and Whites. The relative risks of this more severe hypertension for Whites remaining in each of the bottom two classes, for example, were larger and statistically significant ($P < .02$). The odds ratios for Whites experiencing downward moves into the bottom two classes were also significant for this measure, whereas none of the odds ratios on upward moves among Whites was significant.

Results on the measure of controlled or uncontrolled hypertension are similar to those for uncontrolled hypertension (Table 3). However, the higher odds ratios for Blacks moving into the top class and for those remaining there suggest that these groups may have had better access to medication, adhered more strictly to a medication regimen, or experienced greater efficacy from such medication than did their counterparts experiencing other transitions.

Discussion

Our findings indicate that the incidence of high blood pressure was significantly associated with occupational class position and transition among both Black and White men. The relative risk of hypertension was generally higher among Blacks and among occupational cells near the bottom of the occupational hierarchy where Blacks are most highly concentrated (stayed in class 3 or 4, or moved up to class 3).

The absence of a significant difference among Blacks remaining in the top class calls into question a general relative increase in hypertension by race. Such a conclusion is provocative but preliminary, given the small number of Blacks in this cell. The significant racial differences in hypertension by other occupational class transitions may be related to differences in jobs held by Blacks versus Whites within a given class. An inspection of three-digit census occupations held by race among those remaining in class 2, where the most profound difference in hypertension by race surfaced, failed to indicate obvious differences in occupational patterns for this particular transition. The fact that Whites in the sample more readily experienced upward transitions than did Blacks, however, suggests that the prospect for advancement may be an implicit characteristic that differentiates jobs by race but is not sufficiently captured by the segmentation scheme.

It is also possible that there are unmeasured characteristics of individuals within occupations that could also be responsible for the differences in incidence of hypertension by race. Lower educational level among Blacks, for example, could be a risk factor in hypertension because of education's relationship to higher levels of active coping and because education itself constitutes a resource for coping.[7] The incorporation of measures of education into our analysis, however, did not significantly alter our reported results.

The period covered by this study is widely recognized as one of intense industrial and occupational restructuring in the United States.[28-32] Some of the noted features of this restructuring are captured to some degree as vertical moves from class 2 or 3 to class 4. But several economists have noted that another fundamental aspect of industrial restructuring

during the period, particularly within the craft and manufacturing sectors, has been horizontal segmentation—the downgrading through lower pay, benefits, or job security of what were heretofore "good" jobs, due in part to the intensification of international competition and the decline in union strength in these sectors.[28,30,31] Such horizontal transformations may not be captured in our occupational class matrix. Yet some of the significant results in the regressions for remaining or moving into or out of classes 2 and 3, particularly among Blacks, could reflect such changes.

Wilson has also cited the loss of unskilled and semiskilled manufacturing jobs as a major contributor to the intensification of poverty among Blacks over the period of our analysis, especially in urban centers.[32] The significant results for remaining in or moving into class 4 occupations, highest among Blacks, are likely to capture some effects related to the Wilson hypothesis. We incorporated measures of poverty into our regressions (not shown) to test the extent to which the uncovered differences in hypertension by occupational class and race might be attributable to extreme financial hardship. None of these measures, however, proved to be significant.

Comparative results on the two measures of hypertension suggest that access to, usage of, or the efficacy of antihypertensive medication might be a factor in the uncovered racial differentials in hypertension by occupational class. The precise role that such access or use might play needs to be examined more closely with richer data than were available for this study.

Our analysis is based on blood pressure readings taken at two points in time that spanned several years. While more blood pressure measurements would have been useful, we feel that this cohort study is an improvement over the largely cross-sectional literature relating blood pressure to socioeconomic status.[3-9] To the extent that blood pressure measures at baseline already reflect the effects of occupational class position and transition (see Table 2), our results may understate the full impact of such effects, not only because we adjust for baseline measures of blood pressure but also because we control for other factors related to blood pressure, such as obesity, that have also exhibited social class gradients in other studies.[10,33]

Our use of occupational characteristics as a basis for social stratification has several strengths. The isolation of specific job attributes, such as repetitious tasks and the absence of decision-making authority, that underlie our stratification scheme as potential risk factors for high blood pressure provides a concrete direction for public health policy regarding the design of work. Occupational transitions are also an important basis for changes in income and social status in adult life, whereas other measures of socioeconomic status, such as education, are relatively stationary. Our results relating to both remaining in and moving into jobs where opportunities for training and advancement are typically constrained suggest that reduced prospects for vocational training and occupational mobility may play a formidable role in the incidence of hypertension and may be an important factor in racial differences in hypertension and, by extension, in the incidence of heart disease.

TABLE 3—The Effect of Occupational Class[a] Transitions on the Odds Ratio[b] of Hypertension at Follow-Up

	Whites		Blacks	
	OR	95% CI	OR	95% CI
Equation 1: uncontrolled hypertension[c]				
Stayed in class				
1	1.00	. . .	0.42	0.09, 2.03
2	1.38	0.90, 2.12	21.72	5.63, 83.88
3	1.64	1.13, 2.38	3.98	1.59, 9.99
4	1.46	0.98, 2.18	4.07	2.09, 7.94
Moved down to class				
2	0.96	0.49, 1.89	. . .[d]	. . .
3	1.35	0.83, 2.19	1.92	0.34, 10.88
4	1.42	0.92, 2.19	7.30	2.02, 26.35
Moved up to class				
1	1.76	1.18, 2.65	0.28	0.02, 3.34
2	2.08	1.12, 3.88	1.72	0.26, 11.39
3	1.11	0.51, 2.38	16.08	3.47, 74.50
Model chi-square 364.66; n = 1982; no. of hypertensives = 560.				
Equation 2: controlled or uncontrolled hypertension[e]				
Stayed in class				
1	1.00	. . .	2.55	0.75, 8.69
2	1.25	0.82, 1.89	16.31	4.18, 63.64
3	1.87	1.30, 2.68	4.91	1.83, 13.16
4	1.30	0.88, 1.92	4.62	2.33, 9.15
Moved down to class				
2	0.75	0.39, 1.45	. . .[d]	. . .
3	1.64	1.04, 2.60	1.67	0.29, 9.62
4	1.33	0.87, 2.03	9.47	2.06, 43.45
Moved up to class				
1	1.67	1.12, 2.47	6.36	0.96, 42.27
2	1.56	0.83, 2.91	1.07	0.16, 7.27
3	0.89	0.42, 1.87	14.57	2.75, 77.32
Model chi-square 513.17; n = 1982; no. of hypertensives = 710.				

Note. OR = odds ratio; CI = confidence interval.

[a]In general, jobs in class 1 rank high in educational requirements and/or decision-making authority; jobs in class 2 rank high with respect to vocational training; jobs in class 3 rank low with respect to vocational training and/or high in repetitive operations; jobs in class 4 rank very low with respect to training requirements.

[b]Odds ratios were derived from logistic regression models that also included age, aging between baseline and follow-up exams, body mass index, alcohol use, urban residence, hypertension, and, for equation 1, use of blood pressure medication at follow-up. All covariate measurements were taken at baseline unless otherwise noted.

[c]Follow-up systolic ≥ 140 mm Hg or diastolic ≥ 90 mm Hg.

[d]Only one case in this category.

[e]Follow-up systolic ≥ 140 mm Hg or diastolic ≥ 90 mm Hg or respondent takes blood pressure medication at follow-up.

To reduce racial disparities in high blood pressure and associated disease and mortality is a prominent public health goal of the nation.[1] While the Public Health Service has specifically singled out the work site as a valuable arena for blood pressure screening, education, referral, and control activities,[1] our findings suggest that certain structural aspects of work also merit attention with respect to these goals. □

Acknowledgments

This study was supported by National Research Service Award THS00026 from the Agency for Health Care Policy and Research.

We would like to thank David Ragland, Joan Bloom, and participants in the Behavioral Factors in Cardiovascular Disease Seminar, all at the University of California, Berkeley; participants in the Writing Seminar at the Institute for Health Policy Studies, University of California, San Francisco; Nancy Krieger, Kaiser Foundation Research Institute; and two anonymous reviewers and the Editor of the Journal for their helpful comments and suggestions on earlier versions of this manuscript.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1990. DHHS publication PHS 91–50212.
2. 1988 Joint National Committee. The 1988 report of the Joint National Committee on Detection, Evaluation and Treatment of High Blood Pressure. *Arch Intern Med.* 1988;148:1023–1038.
3. Blood pressure levels of persons 6–74 years: United States, 1971–1974. *Vital Health Stat [11].* 1977; no. 203. DHEW publication HRA 78–1648.
4. Prineas RJ, Gillum R. US epidemiology of hypertension in Blacks. In: Hall WD, Saunders E, Shulman N, eds. *Hypertension in Blacks: Epidemiology, Pathophysiology, and Treatment.* Chicago, Ill: Year Book Medical Publishers, Inc; 1985:17–36.
5. Syme SL, Oakes TW, Friedman GD, Feldman R, Siegelaub AB, Collen M. Social class and racial differences in blood pressure. *Am J Public Health.* 1974;64:619–620.
6. James SA. Psychosocial and environmental factors in Black hypertension. In: Hall WD, Saunders E, Shulman N, eds. *Hypertension in Blacks: Epidemiology, Pathophysiology, and Treatment.* Chicago, Ill: Year Book Medical Publishers, Inc; 1985:132–143.
7. James SA, Hartnett SA, Kalsbeek WB. John Henryism and blood pressure differences among black men. *J Behav Med.* 1983;6:259–278.
8. James SA, La Croix AZ, Kleinbaum DG, Strogatz DS. John Henryism and blood pressure differences among Black men: II. the role of occupational stressors. *J Behav Med.* 1984;7:259–275.
9. James SA, Strogatz DS, Wing SB, Ramsey DL. Socioeconomic status, John Henryism, and hypertension in blacks and whites. *Am J Epidemiol.* 1987;126:664–673.
10. Marmot M, Theorell T. Social class and cardiovascular disease: the contribution of work. *Int J Health Serv.* 1988;18:659–674.
11. Karasek RA, Baker D, Marxer F, Ahlbom A, Theorell T. Job decision latitude, job demands and cardiovascular disease: a prospective study of Swedish men. *Am J Public Health.* 1981;71:694–705.
12. Schnall PL, Pieper C, Schwartz JE, et al. The relationship between "job strain," workplace diastolic blood pressure, and left ventricular mass index. *JAMA.* 1990;263:1929–1935.
13. Theorell T, Perski A, Akerstedt T, et al. Changes in job strain in relation to changes in physiological state: a longitudinal study. *Scand J Work Environ Health.* 1988;14:189–196.
14. Knox SS, Theorell T, Svensson J, Waller D. The relation of social support and working environment to medical variables associated with elevated blood pressure in young males: a structural model. *Soc Sci Med.* 1985;21:525–531.
15. Rumberger RW, Carnoy M. Segmentation in the US labour market: its effects on the mobility and earnings of Whites and Blacks. *Cambridge J Econ.* 1980;4:117–132.
16. Berk RA. An introduction to sample selection bias in sociological data. *Am Sociol Rev.* 1983;48:386–398.
17. Heckman JJ. Sample selection bias as a specification error. *Econometrica.* 1979;45:153–161.
18. Anthropometric reference data and prevalence of overweight: United States, 1976–80. *Vital Health Stat [11].* 1987; no. 238:6. DHHS publication PHS 87–1688.
19. Schatzkin A, Jones Y, Hoover RN, et al. Alcohol consumption and breast cancer in the epidemiologic follow-up study of the first National Health and Nutrition Examination Survey. *N Engl J Med.* 1987;316:1169–1173.
20. Scoville J. The development and relevance of US occupational data. *Ind Labor Relations Rev.* 1965;19:70–79.
21. Scoville JG. *The Job Content of the US Economy, 1940–1970.* New York, NY: McGraw-Hill Book Co; 1969:1–10.
22. Doeringer PB, Piore MJ. *Internal Labor Markets and Manpower Analysis.* Armonk, NY: ME Sharpe Inc; 1971.
23. Edwards R. *Contested Terrain: The Transformation of the Workplace in the Twentieth Century.* New York, NY: Basic Books Inc; 1979.
24. Gordon DM, Edwards R, Reich M. *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States.* New York, NY: Cambridge University Press; 1982.
25. *Dictionary of Occupational Titles.* 4th ed. Washington, DC: US Dept of Labor; 1977.
26. Feinberg SE. *The Analysis of Cross-Classified Categorical Data.* Cambridge, Mass: MIT Press; 1980.
27. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons Inc; 1989.
28. Bluestone B, Harrison B. *The Great U-Turn: Corporate Restructuring and the Polarizing of America.* New York, NY: Basic Books Inc; 1988.
29. Burtless G, ed. *A Future of Lousy Jobs? The Changing Structure of US Wages.* Washington, DC: The Brookings Institution; 1990.
30. Osterman P. *Employment Futures: Reorganization, Dislocation, and Public Policy.* Oxford, England: Oxford University Press; 1988.
31. Pfeffer J, Baron J. Taking the workers back out: recent trends in the structuring of employment. *Res Organ Behav.* 1988;10:257–303.
32. Wilson WJ. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.* Chicago, Ill: University of Chicago Press; 1987.
33. Syme SL, Berkman LF. Social class, susceptibility and sickness. *Am J Epidemiol.* 1976;104:1–8.

Diagnoses, Symptoms, and Attribution of Symptoms among Black and White Inpatients Admitted for Coronary Heart Disease

James M. Raczynski, PhD, Herman Taylor, MD, Gary Cutter, PhD, Michael Hardin, PhD, Neil Rappaport, PhD, and Albert Oberman, MD, MPH

ABSTRACT

Objectives. This study examined health care–seeking behaviors to elucidate factors that contribute to differences in patterns of coronary heart disease between African Americans and Whites. The prevalence of diagnosed coronary heart disease, patients' perceptions of symptoms and attribution of symptoms, and predictors of painful symptoms and attribution of cardiac symptoms were examined.

Methods. The study involved 2416 patients admitted with diagnoses of coronary artery disease, ischemic heart disease, or myocardial infarction or to rule out myocardial infarction. Structured interview questions were used to obtain demographic information, symptoms precipitating admission, and patients' attribution of their symptoms. Discharge diagnoses were obtained from hospital records.

Results. Acute myocardial infarction, unstable angina, nonacute ischemic heart disease, and atherosclerosis were more frequent in White patients. For Blacks, the odds of reporting painful symptoms were only 64% of the odds found for Whites when other factors were controlled, and the odds of attributing symptoms to cardiac origins were almost 50% lower for Blacks than for Whites.

Conclusions. The tendency of Blacks to report fewer painful symptoms and to attribute their symptoms to noncardiac origins may contribute to differences in care-seeking and in medical management of heart disease in Blacks. (*Am J Public Health.* 1994;84:951–956)

Introduction

Coronary heart disease mortality rates for Blacks exceed those of Whites for persons aged 25 through 64 years.[1–4] These differences have led working groups[5] and others[2,3] to call for investigations of factors affecting coronary heart disease in Blacks, particularly individual and environmental factors related to patterns of coronary heart disease that affect care-seeking behavior.[6–9] Prominent among factors hypothesized to affect health care–seeking behavior are the patient's perception of symptoms and attributions made concerning the cause of these symptoms.[10]

As part of a larger project that addresses differences between Black and White patients and community respondents in seeking health care for coronary heart disease (the Birmingham–Black Health Seeking for Coronary Heart Disease Project), we interviewed patients who had been admitted to the hospital for known or suspected cardiac events. The purpose of this paper is to describe the symptoms that were reported by these patients, the frequency of coronary heart disease diagnoses, and relationships between these symptoms and patients' attribution of symptoms.

Methods

Patient Selection

In 1989 and 1990, hospital inpatients were interviewed in three counties in Alabama: Jefferson County, within which Birmingham is located; St. Clair County; and Talladega County. Patients were recruited from five hospitals located in these counties: the University of Alabama Hospital, a large teaching hospital in Birmingham; the Department of Veterans Affairs Medical Center in Birmingham; Cooper Green Hospital, a county hospital for the medically indigent in Birmingham; St. Clair County Hospital, a private hospital serving a rural county; and Coosa Valley Hospital, a private hospital located in Talledega County, also a rural county. Patients approached for interview were those diagnosed with coronary artery disease, ischemic heart disease, chest pain, or myocardial infarction or those admitted to rule out myocardial infarction. Only patients who identified themselves as White or Black completed interviews. Of the patients who were approached, 31 (1.3% of the final sample) refused to be interviewed, and 156 patients terminated the interview prior to completion for various reasons. Of the patients approached, 92.3% completed the interview. The number of participants interviewed was as follows: 1140 White men, 347 Black men, 574 White women, and 355 Black women.

Interview and Data Collection Methods

All interviews were conducted by trained and certified interviewers who underwent periodic retraining to ensure consistency in interview methods. Re-

James M. Raczynski, Herman Taylor, Michael Hardin, Neil Rappaport, and Albert Oberman are with the School of Medicine, University of Alabama at Birmingham. Dr Raczynski is also with the School of Public Health. Gary Cutter is with Pythagoras, Inc, Birmingham, Ala.

Requests for reprints should be sent to James M. Raczynski, PhD, Behavioral Medicine Unit, Division of General and Preventive Medicine, UAB, 101 Medical Towers Bldg, 1717 11th Ave South, Birmingham, AL 35205.

This paper was accepted September 20, 1993.

TABLE 1—Characteristics of Entire Inpatient Sample

	Men			Women		
	White (n = 1140)	Black (n = 347)	*P*	White (n = 574)	Black (n = 355)	*P*
Mean age, y (SE)	60.4 (11.0)	57.5 (12.9)	< .001	63.4 (12.5)	61.1 (13.9)	.032
Mean education, y (SE)	11.5 (3.7)	10.1 (5.9)	< .001	10.6 (3.2)	9.5 (3.5)	< .001
Married/living as married, %	83.6	53.5	< .001	52.2	22.6	< .001
Working full or part time, %	37.5	37.1	.888	22.3	21.1	.681
Hypertensive, %	18.4	35.6	< .001	17.8	39.1	< .001
Smokers, %	66.2	54.9	< .001	72.9	76.8	.189
Hypercholesterolemic, %	32.4	21.8	< .001	35.6	26.6	.005
Diabetic, %	3.7	2.9	.476	5.4	2.3	.023
Positive family history of CHD, %	69.3	44.0	< .001	77.9	62.0	< .001
Personal history of CHD diagnosis, %	66.3	49.4	< .001	56.2	57.9	.617
Particular usual care facility, %	88.4	79.9	< .001	89.9	81.1	< .001
Privately insured, %	44.1	35.1	.003	29.0	17.1	< .001
Medicaid insured, %	3.0	10.1	< .001	10.4	24.3	< .001
Medicare insured %	32.2	26.7	.054	31.7	35.1	.277

Note. CHD = coronary heart disease.

search interviews were conducted in the rooms of willing hospital inpatients. The interviews asked for demographic information, the symptoms that precipitated admission, and attributions for these symptoms at the time they occurred. In addition to the structured interview, discharge diagnoses were obtained from hospital medical records.

Statistical Analyses

Participants were stratified by sex and initial comparisons were conducted for key demographic characteristics between races. Chi-square analyses were used to address differences between racial groups in the frequency distributions obtained from demographic and other key questionnaire items. Frequencies of diagnostic category groupings were compared for racial groups. Perceived symptoms and attributions of symptoms were then examined for differences between Blacks and Whites by sex with chi-square analyses. To ensure that results were similar for the entire patient sample and patients with definite coronary heart disease (*International Classification of Diseases*, ninth revision[11] [ICD-9] diagnoses 410–414), analyses were completed for both sets of patients. Because chest pain is the primary symptom associated with coronary heart disease and either chest pain or pain in other areas might motivate patients to seek treatment, we were interested in assessing the factors that were associated with painful symptoms and attribution of symptoms to cardiac origins. Attribution of symptoms was thus examined for the entire patient sample. Multiple logistic regression analyses were used to examine the relationship of selected demographic characteristics to painful symptoms and attribution of symptoms to cardiac origins. Like the symptom and attribution analyses, these multiple logistic regression analyses were completed both for the entire patient sample and separately for the patients with coronary heart disease (ICD-9 diagnoses 410–414) and those in other diagnostic groups.

Results

Participant Characteristics

Characteristics of respondents are summarized in Table 1. A total of 1487 men (1140 Whites and 347 Blacks) and 929 women (574 Whites and 355 Blacks) were interviewed. Of the entire sample, 29.1% were Black and 38.5% were women. While the ratio of White male inpatients to White female inpatients was almost 2 to 1, the numbers of black Male and Female inpatients were comparable. White participants were older and more educated than their Black counterparts. Fewer Blacks than Whites were married or living as married. Comparable numbers of Whites and Blacks within each sex group reported working full or part time. Differences between the racial groups emerged in self-reported risk factors. Hypertension was reported almost twice as often by Black respondents as by Whites. A higher percentage of White than Black men smoked, and more Whites than Blacks of both sexes reported being hypercholesterolemic. Diabetes was reported very infrequently in all groups. More White than Black respondents reported a positive family history of heart disease. Comparable percentages of Black and White women reported having a history of a coronary heart disease diagnosis, but a lower percentage of Black men than White men reported a previous coronary heart disease diagnosis. Although more than 80% of the entire sample reported having a usual source of care, more Whites than Blacks reported having a usual care source. A higher percentage of White than Black respondents were covered by private insurance, although less than half of the entire sample (less than a third for women) was privately insured. Medicaid coverage was greater among both Black men and women than among Whites. However, despite the Medicare and Medicaid coverage, approximately twice as many Blacks as Whites were uninsured.

Frequency of Coronary Heart Disease Diagnoses

Table 2 summarizes the frequency of discharge diagnoses by sex and racial groups. White men and women were more likely than their Black counterparts to have diagnoses of acute myocardial infarction and unstable angina (ICD-9 diagnoses 410–411) and nonacute ischemic heart disease and atherosclerosis (ICD-9 diagnoses 412–414). Black men and women were more likely than Whites to have other forms of heart disease (ICD-9 diagnoses 420–429) and to not have any cardiac diagnosis.

Symptom Perception and Attribution

Table 3 describes the symptoms that precipitated hospital admission for the entire sample of patients and for those patients who had diagnoses of myocardial infarction, ischemic heart disease, or

atherosclerosis (ICD-9 diagnoses 410–414). In the entire sample, Whites reported more arm pain and numbness than Blacks. When all painful symptom categories were collapsed to examine patients who reported any painful symptom, White men were found to report more overall painful symptoms than Black men (75% vs 68%). The data for women showed similar trends (76% vs 70%), suggesting that White women also experienced more painful symptoms than their Black counterparts. The number of patients reporting any nonpainful symptom, obtained from collapsing across all nonpainful symptom categories, revealed that Blacks reported more nonpainful symptoms than Whites. Restricting the analysis to patients with coronary heart disease diagnoses revealed somewhat different results. More than twice as many White men as Black men reported arm pain and numbness. There were no differences in painful symptom reports between racial groups for women, and neither sex showed differences between racial groups when all of the painful symptom categories were collapsed. Black female coronary heart disease patients reported more overall nonpainful symptoms than did White female coronary heart disease patients, as was true for the entire patient sample, but men reported no overall differences in nonpainful symptoms. Very few patients reported no symptoms associated with their admission.

Table 4 displays the initial attributions that patients made for their symptoms. About one quarter of the White men in the entire sample attributed their symptoms to heart attack and one quarter to angina, and these proportions remained when only the coronary heart disease patients were considered. However, fewer Black men in the entire patient sample attributed their symptoms to heart attack, and fewer Black men in both patient groups thought that their symptoms were due to angina. In the entire patient sample, but not in the coronary heart disease patient sample, Black men were more likely than Whites to attribute symptoms to causes that are not listed in Table 4. For women in the entire patient sample, differences between racial groups in attribution of symptoms to myocardial infarction and angina were similar to those found for the men. However, Black and White women in the coronary heart disease sample did not differ in their attribution of symptoms to myocardial infarction and angina.

TABLE 2—Frequency of Discharge Diagnoses, by Race and Sex

Diagnosis	Men			Women		
	White	Black	*P*	White	Black	*P*
Acute myocardial infarction and unstable angina,[a] %	41.8	31.2	.17	48.5	26.0	<.001
Nonacute ischemic heart disease and atherosclerosis,[b] %	79.4	48.2	<.001	69.9	40.8	<.001
Heart disease, other forms,[c] %	48.5	60.1	<.001	58.0	76.4	<.001
No cardiac diagnosis, %	9.3	22.3	<.001	9.6	21.6	<.001

Note. Categories are not exclusive; therefore, columns add to more than 100%.
[a]*International Classification of Diseases*, Ninth Revision (ICD-9), codes 410–411.
[b]ICD-9 codes 412–414.
[c]ICD-9 codes 420–429.

TABLE 3—Percentage Distribution of Symptoms Precipitating Admission, by Sex and Race

Symptom	Men			Women		
	White	Black	*P*	White	Black	*P*
Entire patient sample						
Chest pain/discomfort	70.7	67.3	.242	72.8	67.4	.084
Arm pain/numbness	19.8	9.5	<.001	22.7	15.4	.008
Back pain	3.0	3.5	.634	6.5	3.4	.046
Neck/jaw/ear pain	3.3	1.5	.068	4.7	2.9	.171
Any painful symptoms	74.6	67.9	.015	75.7	70.3	.073
Sweating	8.9	10.7	.312	6.7	10.6	.037
Unconsciousness	1.9	3.8	.044	1.3	3.1	.048
Shortness of breath	26.1	29.2	.261	29.7	38.9	.005
Nausea/vomiting	8.5	11.0	.166	11.2	14.0	.206
Indigestion/fullness	3.3	6.4	.012	3.4	6.0	.066
Other symptoms	30.0	32.1	.459	30.3	32.9	.414
Any nonpainful symptoms	48.6	57.5	.004	51.5	65.7	<.001
No symptoms	2.2	3.5	.169	2.9	1.4	.157
Coronary heart disease patients[a]						
Chest pain/discomfort	74.1	78.9	.252	75.0	76.3	.795
Arm pain/numbness	22.1	10.2	.002	24.3	17.2	.159
Back pain	2.5	3.9	.370	6.6	8.6	.521
Neck/jaw/ear pain	3.4	0.8	.155	5.5	4.3	.791
Any painful symptoms	78.2	78.9	.866	77.9	79.6	.742
Sweating	8.5	15.6	.013	7.0	11.8	.142
Unconsciousness	1.2	1.6	.664	0.0	1.1	.255
Shortness of breath	24.3	23.4	.841	23.2	30.1	.181
Nausea/vomiting	8.2	16.4	.054	13.2	15.1	.660
Indigestion/fullness	3.7	7.8	.035	3.7	8.6	.058
Other symptoms	26.9	19.5	.080	26.1	22.6	.499
Any nonpainful symptoms	46.8	51.6	.319	45.2	57.0	.050
No symptoms	2.2	4.7	.105	2.6	1.1	.685

Note. Categories are not exclusive; therefore, columns add to more than 100%.
[a]*International Classification of Diseases*, Ninth Revision, codes 410–414.

We also examined separately the attributions made by those patients who reported having chest pain as a symptom, and the findings were very similar to those for the entire patient sample. Black patients were less likely than Whites to attribute their chest pain to cardiac origins (40.9% vs 61.0% for Black and White men, respectively, $P < .001$; 48.7% vs 57.7% for Black and White women, respectively, $P = .028$). Attribution of chest pain symptoms to indigestion, heartburn, or stomach or esophageal problems was also more common among Black than

TABLE 4—Attribution of Symptoms (%), by Sex and Race

Attribution	Men			Women		
	White	Black	*P*	White	Black	*P*
Entire patient sample						
Heart attack or MI	27.5	18.7	.001	21.9	21.7	.936
Angina or heart pain	24.2	12.6	<.001	26.3	17.1	.001
Cardiac origin[a]	51.8	31.3	<.001	48.2	38.9	.006
Flu or other illness	1.8	2.9	.215	2.2	2.9	.506
Indigestion/heartburn, stomach or esophageal problem	11.7	18.7	.001	10.3	15.1	.028
Muscle pain	1.4	0.9	.588	1.6	0.6	.219
Breathing or lung problem	2.1	3.7	.079	2.2	6.0	.003
Liver problem	0	0	. . .	0	0	. . .
Gall bladder problem	0.2	0	. . .	0.4	0.3	.999
Other	24.4	34.8	<.001	29.3	30.6	.688
Coronary heart disease patients[b]						
Heart attack or MI	27.7	22.7	.243	23.2	26.9	.469
Angina or heart pain	24.1	11.7	.002	26.5	25.8	.900
Cardiac origin[a]	51.8	34.4	<.001	49.6	52.7	.611
Flu or other illness	1.5	3.1	.255	0.7	0	. . .
Indigestion/heartburn, stomach or esophageal problem	13.4	27.3	<.001	11.0	17.2	.121
Muscle pain	1.6	0.8	.703	1.5	1.1	.999
Breathing or lung problem	2.1	0.8	.487	1.8	1.1	.999
Liver problem	0	0	. . .	0	0	. . .
Gall bladder problem	1.3	0	. . .	0.4	0	. . .
Other	23.8	27.3	.395	29.0	23.7	.316

[a]Reflects attribution to either myocardial infarction (MI) or angina.
[b]*International Classification of Diseases*, Ninth Revision, codes 410–414.

TABLE 5—Summary of Multiple Logistic Regression Analyses for Predictors of Painful Symptoms and Attribution of Symptoms to Cardiac Origins, by Sex and Race, Entire Patient Sample

Variable	Painful Symptoms		Attribution to Cardiac Origins	
	OR	95% CI	OR	95% CI
Race: Black	0.635*	0.506, 0.798	0.548**	0.441, 0.681
Sex: female	1.058	0.863, 1.296	1.086	0.893, 1.320
CHD discharge diagnosis[a]	1.487*	1.236, 1.790	1.231**	1.030, 1.473
Age	1.009	0.999, 1.019	0.997	0.988, 1.007
Education	1.014	0.991, 1.038	1.029**	1.004, 1.053
Married status	1.131	0.913, 1.400	0.983	0.801, 1.206
Employment	1.202	0.945, 1.529	0.912	0.722, 1.151
Hypertension	1.078	0.855, 1.358	1.213	0.969, 1.518
Smoking	1.230*	1.002, 1.510	1.164	0.954, 1.418
Hypercholesterolemia	1.017	0.819, 1.262	1.112	0.904, 1.367
Diabetes	0.878	0.540, 1.427	0.830	0.522, 1.318
Positive family history of CHD	0.708*	0.582, 0.862	1.024	0.846, 1.240
Personal history of CHD diagnosis	0.927	0.769, 1.118	2.522**	2.100, 3.029
Particular usual care facility	1.952*	1.610, 2.365	1.381**	1.143, 1.668
Private insurance	1.360**	1.051, 1.759	0.912	0.713, 1.165
Medicare insurance	1.296**	1.001, 1.678	1.137	0.891, 1.451
Medicaid insurance	1.287	0.883, 1.873	0.881	0.622, 1.250

Note. OR = odds ratio; CI = confidence interval; CHD = coronary heart disease.
[a]*International Classification of Diseases*, Ninth Revision, codes 410–414.
*$P < .001$; **$P < .05$.

White patients (22.6% vs 13.6% for Black and White men, respectively, $P = .001$; 18.6% vs 11.4% for Black and White women, respectively, $P = .011$).

Predictors of Chest Pain and Attribution of Symptoms to Cardiac Origins

Results of the multiple logistic regression analyses for predictors of painful symptoms and attribution of symptoms to cardiac origins are summarized in Table 5. Factors associated with the report of painful symptoms were being White, having a coronary heart disease diagnosis, smoking, not having a positive family history of coronary heart disease, having a particular usual care facility, having private insurance, and having Medicare insurance. To examine further the influence of diagnosis on painful symptom prediction, we conducted additional multiple logistic regression analyses with different diagnostic group classifications (ICD-9 diagnoses 410–411 and 410–429) coded in the coronary heart disease diagnosis variable of our model. The results of these analyses confirm the significant independent contributions of race and having a particular usual care facility.

Factors associated with attributing symptoms to cardiac origins were being White, having a coronary heart disease diagnosis, being more educated, having previously been diagnosed as having coronary heart disease, and having a particular usual care facility. Repeated analyses with our different diagnostic group classifications (ICD-9 diagnoses 410–411 and 410–429) coded in the coronary heart disease diagnosis variable of our model similarly suggested that race is an independent predictor of attribution of symptoms to coronary heart disease origins.

Discussion

The experience of painful symptoms and the attribution of symptoms to cardiac origins are two factors that presumably affect treatment seeking and possibly influence medical management. As in previous studies,[12,13] the African-American patients in our sample were younger than the White patients. Self-reported risk factor differences emerged between Whites and Blacks in our sample. Black men and women were twice as likely as their White counterparts to be hypertensive. Whites reported higher rates of hypercholesterolemia and family history of heart disease. Differences in access to

care also emerged between racial groups. More Whites than Blacks reported that they had a particular source of usual care. The economic advantage of the White patients was also reflected in rates of insurance coverage.

We found that Whites were more likely to have discharge diagnoses of acute myocardial infarction and unstable angina (ICD-9 diagnoses 410–411) and nonacute ischemic heart disease and atherosclerosis (ICD-9 diagnoses 412–414) and that Blacks were more likely to have discharge diagnoses of other forms of heart disease (ICD-9 diagnoses 420–429) and not to have cardiac diagnoses. This finding that coronary heart disease is less common among Blacks than whites, at least among those seen as inpatients, is similar to findings from the Coronary Artery Surgery Study[14] and the Myocardial Infarction Triage and Intervention study.[12]

Symptoms associated with hospital admission also differed between Whites and Blacks. In our overall patient sample, Whites of both sexes reported experiencing slightly more arm pain and numbness than did Blacks. Overall, White men reported more painful symptoms than Black men. More nonpainful symptoms were reported by Blacks than Whites of both sexes. Thus, the symptom patterns of Whites and Blacks seem to differ in a sample of patients admitted for suspected coronary heart disease. Although similar percentages of Whites and Blacks within each sex group experienced the primary symptom of heart disease, that is, chest pain or discomfort, the overall symptom pattern suggests that more nonpainful symptoms may be reported by Blacks than by Whites. The effect of these nonpainful symptom reports on medical decision making is not addressed by our data; however, one might speculate that patients' reports of painful symptoms would influence medical decision making to at least some degree. Conversely, the lack of classic symptoms can also affect medical decision making and the diagnostic workup.

Restricting our symptom analysis to patients with acute diagnoses of coronary heart disease (ICD-9 diagnoses 410–414) produced somewhat different results. Generally, both Blacks and Whites with coronary heart disease reported similar symptoms, although some differences persisted. Significantly fewer painful symptoms and more nonpainful symptoms were reported by Black patients.

In addition to differences between Whites and Blacks in the perception of symptoms, we found differences in attribution of symptoms. Approximately half of the White men and women in our overall patient sample had attributed their symptoms to cardiac origins, whereas fewer than one third of Black men and slightly more than one third of Black women had thought that their symptoms were related to their hearts. Correspondingly, Blacks were more likely to attribute their symptoms to a host of other origins. Restricting our analysis to patients with acute coronary heart disease diagnoses revealed similar differences in attribution for male patients; for women, however, the pattern of symptom attribution was similar for Blacks and Whites. Analyses of data for only those patients who reported chest pain showed that again, Blacks less commonly attributed their symptoms to cardiac origins and more commonly attributed them to noncardiac origins. These differences in attribution may influence the care-seeking behavior of patients, which, in turn, may influence medical decision making that is associated with patients' perceptions of their symptoms. Since the patients we studied are the patients who made it to the hospital, several key questions remain. In addition to the selective nature of our inpatient sample, selective attrition due to factors such as sudden death may have biased our results.

As the results of the multiple logistic regression analyses emphasize, race was independently associated with both perception of painful symptoms and attribution of symptoms to cardiac origins. For Blacks, the odds of reporting painful symptoms were only 64% of the odds found for Whites when other factors were controlled, and the odds of attributing symptoms to cardiac origins were almost 50% lower for Blacks than for Whites. It is possible that ethnic background may account for differences in symptom perception and attribution in a number of different ways. First, it may be that there are physiological mechanisms that result in lower rates of perceived pain. Second, cultural and ethnic background may be associated with pain tolerance.[13] Berkanovic and Telesky suggest that a historical stoic response to illness may persist among Blacks.[15] This and other cultural factors may affect the perception of symptoms.[16] Finally, reports indicate that Blacks are less knowledgeable than Whites about heart disease symptoms,[17] suggesting that they may be less aware of heart disease, less alert to painful symptoms, and less likely to attribute symptoms to cardiac origins. Alternatively, these reports may reflect differences between Blacks and Whites in the presentation of cardiac disease.

In addition to race, our multiple logistic regression analyses found other variables that were associated with painful symptoms. The association of diagnosis with painful symptom reports is as might be predicted: patients with coronary heart disease discharge diagnoses were more likely to report painful symptoms. Several other relationships also emerged, although these relationships were not consistent across models. Smoking was the only risk factor associated with painful symptom reports, and having a family history of heart disease was associated with not having painful symptoms. Having a particular usual care facility was associated with having painful symptoms in all of the multiple logistic regression analyses. Whether or not a person has a particular usual care facility would not seem to directly affect the experience of cardiac symptoms, although it is possible that persons who have more overt symptoms, including pain, are more likely to seek out a usual care provider. Alternatively and perhaps more likely, it may be that persons who seek primary care are those who do not deny their symptoms and are therefore more likely to report painful symptoms when they occur. Patients with private insurance and Medicare insurance were more likely than their uninsured counterparts to have painful symptoms, again possibly reflecting either the seeking of insurance coverage for those experiencing pain or less denial about pain among those privately insured.

Like reports of painful symptoms, diagnosis was associated with attribution of symptoms to cardiac origins. As one might expect, patients with coronary heart disease diagnoses, whether at discharge or prior to admission, were more likely to attribute symptoms to cardiac origins. Patients with more education were more likely to attribute their symptoms to cardiac origins, perhaps reflecting more knowledge about what symptoms are associated with coronary heart disease or a greater willingness to accept the possibility that they have coronary heart disease. Having a particular usual care facility was also associated with attribution of symptoms, perhaps for the same reasons that it is related to reporting painful symptoms.

It should be noted that there are several limitations in our data and the inferences that can be made from them. The methods that we used resulted in our

examining inpatients who not only had made it to treatment but who were admitted and then stabilized, thus excluding those who never made it to treatment or who were turned away from treatment. The cross-sectional design also limits the causal inferences that can be made. The reliance on self-report data, with their inherent biases, is also a limitation. Severity of symptoms and comorbidity may be related to symptom patterns and attribution, but these factors were not examined. Finally, these data were obtained largely from patients who reside in Alabama, and the generalizability of the findings to patients from other regions may be limited.

Despite these limitations, our results suggest that among our sample of patients who sought and were admitted for treatment, race was independently associated with both perception of painful symptoms and attribution of symptoms to cardiac origins. When factors other than race were controlled, African-American patients were less likely than Whites to report either experiencing painful symptoms or attributing their symptoms to cardiac origins. While cultural factors and physiological mechanisms may account for these differences in perception, knowledge about symptoms of coronary heart disease may also account for our findings. Lowered awareness of coronary heart disease and its symptoms might be expected to lead to less alertness to painful symptoms and less likelihood of attributing symptoms to cardiac origins. If this proves to be the case, as has been suggested by preliminary reports,[18,19] then educational interventions may be indicated for overcoming these discrepancies. In addition, these findings may suggest that health care providers should be educated about racial differences in symptom perception and attribution. □

Acknowledgments

This work was supported by grant HL39197 from the National Heart, Lung, and Blood Institute.

Preliminary findings presented in this paper were presented at the Fourth National Forum on Cardiovascular Health, Pulmonary Disorders, and Blood Resources (Minority Health Issues for an Emerging Majority), Washington, DC, June 1992.

The authors wish to acknowledge the assistance of Jay Hsu in conducting the programming for the statistical analyses reported in this paper.

References

1. Gillum RF. Coronary heart disease in black populations, I: mortality and morbidity. *Am Heart J.* 1982;104:839.
2. Watkins LO. Epidemiology of coronary heart disease in black populations: methodologic proposals. *Am Heart J.* 1984;108: 635–640.
3. *Report of the Secretary's Task Force on Black and Minority Health.* Washington, DC: US Dept of Health and Human Services; 1985. GPO publication 017-090-00078-0.
4. Castaner A, Simmons BE, Mar M, Cooper R. Myocardial infarction among black patients: poor prognosis after hospital discharge. *Ann Intern Med.* 1988;109:33–35.
5. Johnson KW, Payne GH. Report of an NHLBI working conference on coronary heart disease in black populations: preface. *Am Heart J.* 1984;108:633–634.
6. Adams L, Africano E, Doswell W, et al. Summary of workshop I: working group on epidemiology. *Am Heart J.* 1984;108:699–702.
7. James SA. Socioeconomic influences on coronary heart disease in black populations. *Am Heart J.* 1984;108:669–672.
8. Kasl SV. Social and psychologic factors in the etiology of coronary heart disease in black populations: an exploration of research needs. *Am Heart J.* 1984;108:660–669.
9. Caldwell J, Eaker E, Edozien J, et al. Summary of workshop III: working group on socioeconomic and sociocultural influences. *Am Heart J.* 1984;108:706–710.
10. King JB. Illness attributions and the Health Belief Model. *Health Educ Q.* 1984;10:287–312.
11. *Manual of the International Classification of Diseases, Injuries, and Causes of Death.* 9th revision. Geneva, Switzerland: World Health Organization; 1977.
12. Maynard C, Litwin PE, Martin JS, et al. Characteristics of black patients admitted to coronary care units in metropolitan Seattle: results from the Myocardial Infarction Triage and Intervention Registry (MITI). *Am J Cardiol.* 1991;67:18–23.
13. Strogatz DA. Use of medical care for chest pain: differences between Blacks and Whites. *Am J Public Health.* 1990;80:290–294.
14. Maynard C, Fisher LD, Passamani ER, Pullum R. Blacks in the Coronary Artery Surgery Study: risk factors and coronary artery disease. *Circulation.* 1986;74:64–71.
15. Berkanovic E, Telesky C. Mexican-American, Black-American and White-American differences in reporting illnesses, disability and physician visits for illnesses. *Soc Sci Med.* 1985;20:567–577.
16. Bailey EJ. Sociocultural factors and health care-seeking behavior among Black Americans. *J Natl Med Assoc.* 1987;79:389–392.
17. Folsom AR, Sprafka JM, Luepker RV, Jacobs DR. Beliefs among black and white adults about causes and prevention of cardiovascular disease: the Minnesota Heart Survey. *Am J Prev Med.* 1988;4:121–127.
18. Kannel WB, Abbott BD. Incidence and prognosis of unrecognized myocardial infarction: an update on the Framingham Study. *N Engl J Med.* 1984;311:1144–1147.
19. Lewis CE, Raczynski JM, Oberman A, Cutter GR. Natural history of coronary heart disease in blacks. In: Saunders E, Brest AN, eds. *Cardiovascular Diseases in Blacks.* Philadelphia, Pa: F.A. Davis Co; 1991:29–45.

Do Blacks and Whites Differ in Their Use of Health Care for Symptoms of Coronary Heart Disease?

Sybil L. Crawford, PhD, Sarah A. McGraw, PhD, Kevin W. Smith, MA, John B. McKinlay, PhD, and Judith E. Pierson, MSW

ABSTRACT

Objectives. The purpose of this study was to identify the role of race in seeking and receipt of care for symptoms of coronary heart disease.

Methods. Data on medical care, sociodemographic characteristics, symptoms, risk factors, income, and insurance were collected in a telephone interview for a random sample of 2030 Black and White adults in inner-city Boston. Rates of care-seeking for symptoms, amounts of delay in seeking care, and rates of receipt of care were compared for Blacks and Whites after adjustment for other characteristics.

Results. Before and after adjustment for other factors, Blacks and Whites were equally likely to seek care. Average delay time was shorter for Blacks, particularly Black women. With the exception of a lower rate of referral to cardiologists among Blacks, receipt of care was similar for Blacks and Whites who sought medical attention for symptoms.

Conclusions. In an urban population of Blacks and Whites who were similar in socioeconomic status and access to medical care, there were few racial differences in coronary heart disease–related care patterns. (*Am J Public Health.* 1994;84:957–964)

Introduction

Recent research suggests that Blacks are less likely than Whites to seek and to receive appropriate medical care for coronary heart disease, the leading cause of death in US Blacks.[1,2] In the Edgecombe County Study, nearly half of all Blacks with repeated chest pain never discussed this pain with a doctor, compared with 27% of Whites.[3] In another study, Blacks admitted to a municipal hospital with documented acute myocardial infarction had waited almost twice as long as Whites before seeking medical care.[4] Other research suggests that Blacks make less use of ambulatory care, including office visits for coronary heart disease and other circulatory system diseases.[5–8] Additional work indicates that Blacks seeking medical care related to coronary heart disease receive lower levels of treatment such as coronary angiography or coronary artery bypass grafting, even after disease severity is controlled,[9–14] and that they receive fewer general and coronary heart disease–related diagnostic and treatment services.[15,16]

Differences in coronary heart disease–related health care use may account for the observed similar or lower prevalence of coronary heart disease[17–19] but higher case-fatality rates and excess mortality[19–27] in Blacks than in Whites. Delays in seeking care, which are negatively associated with prognosis and survival,[4,28,29] could contribute to higher mortality among Blacks with coronary heart disease. Alternatively, reported incidence and prevalence rates, which are often derived from hospital discharge records rather than from population-based studies, may reflect underdetection of coronary heart disease among Blacks owing to their lower rates of help-seeking.[1] Lower levels of care received by Blacks could contribute to both underdetection of coronary heart disease and a higher mortality rate.

These hypotheses underscore the importance of population-based samples, since studies based on unrepresentative samples (e.g., hospital discharge records) can miss persons with coronary heart disease who are not in the medical care system, who may be disproportionately Black. Moreover, non–population-based samples may lead to a biased assessment of help-seeking and care patterns. For example, one study of delay based on hospitalized patients excluded subjects who failed to reach the hospital, some of whom may have been sudden deaths who did not delay in seeking medical assistance.[4] In addition, if Blacks who seek care are less likely than Whites to receive assistance, estimates of rates of help-seeking by Blacks based on hospital or other medical records will be too low, and lower levels of receipt of care in Blacks may be attributed incorrectly to failure to seek care. Thus population-based sampling is critical in studying racial differences in various stages of health care patterns.

A complete assessment of racial differences in utilization or receipt of health care involves identifying not only the extent of such differences, but their sources as well. That is, if there are Black–White differences in patterns of

The authors are with the New England Research Institute, Watertown, Mass.

Requests for reprints should be sent to Sybil L. Crawford, PhD, New England Research Institute, 9 Galen St, Watertown, MA 02172.

This paper was accepted September 23, 1993.

care, can they be explained by factors related to race? For example, past studies have found that income and insurance coverage explained racial differences in care received.[15,30] Thus racial differences in coronary heart disease–related care patterns may be due in large part to factors related to race, such as socioeconomic status (SES), rather than to race per se.

This study examined patterns of coronary heart disease-related care in a community-based random sample of Black and White adults with similar levels of SES and geographic access to care. First, racial differences in care-seeking in response to coronary heart disease symptoms, both in whether care was sought and in the amount of delay, were assessed before and after controlling for other factors. Second, among those who sought medical assistance, the extent of unadjusted and adjusted racial differences in care received were identified.

Methods

Subject Selection

Data were from a random-digit-dialed telephone survey conducted in three inner-city Boston neighborhoods. The sample was drawn from the community rather than from hospital discharge records to include persons not in regular contact with the health care system. The study's target area consisted of seven zip codes encompassing the Dorchester, Roxbury, and Mattapan neighborhoods. These neighborhoods were selected for their relative homogeneity of household income and geographic access to care, with four major hospitals within the study boundaries and five teaching hospitals bordering the study area. Study respondents were identified through random-digit dialing in the 11 telephone exchanges serving the target area. The same number of dialings was made in each exchange to ensure that all households with telephones in the area had the same probability of being contacted. Eligibility for the study was restricted to Black and White adults born in the United States who were aged 44 to 75 years at the time of the interview. All eligible adults in a household were asked to complete a telephone interview approximately 20 minutes in length. Interviews were conducted between September 1988 and December 1989.

Outcome Measures

Respondents were asked about occurrence of the following coronary heart disease–related symptoms: chest pain, discomfort, pressure, or heaviness, collectively referred to below as "chest pain," and shortness of breath. For the purposes of this study, subjects who reported *ever* experiencing one or more of these symptoms were considered to be symptomatic. Subjects reporting a symptom were asked if they had ever seen a physician or other health care professional in response to the symptom, and if so, the time between first noticing the symptom and contacting the health care professional. Symptomatic respondents who reported ever contacting a medical person in response to chest pain or shortness of breath were considered to have sought help. Delay in help-seeking, measured separately for chest pain and shortness of breath, was defined as the amount of time (in hours) between the onset of the symptom and contact with a medical person. Receipt of care involved eight types of coronary heart disease–related care in three areas: (1) physician-recommended diagnostic procedures: electrocardiogram, treadmill or bicycle exercise test, and echocardiogram; (2) treatment: physician-recommended invasive procedures (coronary artery bypass grafting, angioplasty, and cardiac catheterization), current cardiovascular medication, and hospitalization for myocardial infarction; and (3) other coronary heart disease–related contacts with the health care system: referral to a cardiologist and diagnosis of a heart problem (large heart, heart disease, heart valve problem or heart murmur as an adult, and angina or hardening of the arteries).

Predictor Measures

Predictor information included the following sociodemographic characteristics: race, sex, age, and living situation (with someone or alone). SES measures included high school completion; amount of difficulty in paying for basic necessities, including health care; and current employment status. Respondents were asked about coverage by private insurance, Medicaid, and Medicare. Coronary heart disease risk factors included diabetes, family history of heart disease or stroke, smoking status (current, stopped in past 10 years, never smoked/stopped over 10 years ago), body mass index, diagnosed hypertension, and diagnosed elevated cholesterol. The severity of coronary heart disease symptoms at the time of first occurrence was also noted. Questions about access to and satisfaction with the medical care system were adapted from items used in previous studies of help-seeking[3,31] and included having a usual source of care other than a hospital emergency room, travel time to usual care source, amount of difficulty in traveling to care, level of satisfaction with the amount of time spent with the doctor, level of satisfaction with care received, and need for care not being received. General propensity to seek care was assessed by asking whether the respondent would seek help for six conditions: swelling of the ankles, chronic fatigue, shortness of breath, fainting spells, chest pain, and persistent coughing.[32] Myocardial infarction symptom knowledge was defined as the number of symptoms mentioned in free recall, including chest pain, arm pain, shortness of breath, sweating, nausea, fatigue, dizziness, fainting, and loss of consciousness.

Statistical Analyses

Predictors of help-seeking in response to chest pain were identified by estimating a multivariate logistic regression model for the probability of seeking care by subjects with chest pain. A corresponding model was estimated for the probability of seeking care for shortness of breath by subjects with shortness of breath. For subjects who sought care for chest pain, a multivariate linear regression model was estimated for the amount of delay in seeking care. A similar model was estimated for delay in help-seeking for shortness of breath. Because hours of delay was highly skewed for both symptoms, a log transformation was applied.[33] For receipt of care by symptomatic persons who sought help, separate logistic regression models were estimated for each of the eight types of care. Unadjusted racial differences for each of the outcomes were assessed by including only race as a predictor. Corresponding adjusted racial differences were obtained by adding the remaining predictors. For ease of interpretation, nonoverlapping insurance categories were defined as any private insurance, any Medicare without private insurance, Medicaid only, and uninsured. Because previous analyses of coronary heart disease incidence[1,34] and help-seeking[3] indicated that racial differences vary by sex, the interaction between race and sex was included as a predictor.

Stepwise and backward elimination procedures[33,35] (with $P \leq .05$) were employed to eliminate redundant or unre-

lated covariates from the multivariate models in order to better estimate the effects of the remaining predictors. Model fit was assessed with the Hosmer-Lemeshow statistic[36] for logistic regression models and residual diagnostics for the linear regressions.[33] Computations were performed with SAS[37] and BMDP[38] software. Potential nonresponse bias was examined by comparing study respondents and nonrespondents with regard to demographic characteristics from the household screening. The effect on standard errors of clustering of respondents within households was assessed by calculating design effects for each outcome.[39]

Results

Participants

Random-digit dialing in the target area exchanges identified 2310 adults meeting the study's eligibility criteria. Eligibility status could not be determined for an additional 545 persons who did not complete an individual screening. Telephone interviews were successfully completed with 1222 Black and 808 White respondents (overall response rate = 87.9% among known eligible persons). Of the eligible persons who were not interviewed, 232 refused to participate and 48 could not be recontacted after the initial screening. Because the composition of the respondent sample (39.8% White, 37.8% male) differed very little from that of the full potential sample of 2310 persons (41.2% White, 39.2% male), no adjustments were made for nonresponse. Estimated design effects were very close to 1.00 (ranging from 0.99 to 1.06), indicating that adjustment for household clustering was not necessary.

Of the respondents, 627 Blacks (51.3%) and 380 Whites (47.0%) reported ever experiencing at least one coronary heart disease symptom. Although household income was generally low, Blacks tended to have lower levels of SES than Whites, as seen in education, difficulty in paying for basic necessities, and insurance coverage (Table 1). Blacks had higher rates of diabetes and hypertension but a lower rate of family history of coronary heart disease. Body mass index was slightly lower on average in Whites. The two groups were similar in perceived seriousness of symptoms. With the exception of longer travel times for Blacks, access and satisfaction with care were similar in the two groups. Blacks had a higher propensity to seek care, whereas Whites' knowledge of myocardial infarction symptoms was slightly higher.

TABLE 1—Characteristics of Symptomatic Respondents

	Blacks (n = 627)	Whites (n = 380)
Demographics		
Female, %	66.2*	54.7
Mean age, y (SD)	56.0 (8.1)*	58.8 (8.7)
Living alone, %	28.0	24.2
Socioeconomic status		
Completed high school, %	56.6*	79.2
Very difficult paying for basics, %	37.3*	18.3
Currently employed, %	53.1	53.4
Insurance		
Private, %	62.2*	79.8
Medicaid, %	21.8*	6.4
Medicare, %	29.5**	35.5
Uninsured, %	11.7***	6.6
Risk factors		
Current smoking, %	36.1	38.4
Diabetes, %	19.4**	14.2
Family history, %	42.2*	52.8
Mean body mass index (SD)	28.7 (6.7)*	26.8 (5.6)
Diagnosed hypertension, %	64.1*	48.2
Diagnosed elevated cholesterol, %	30.9	34.4
Symptoms		
No chest pain, %	22.1	24.0
Somewhat/very serious chest pain, %	43.4	40.9
No SOB, %	40.5	37.9
Somewhat/very serious SOB, %	29.8	30.8
Access/satisfaction		
> 30 min travel time, %	18.9*	10.0
Enough time with physician/nurse, %	93.5	93.0
No difficulty in getting to care, %	83.5	85.5
Usual source of care %	86.6	85.0
Very satisfied with care, %	77.7	81.2
Not getting needed care	21.7***	14.8
Propensity/knowledge		
Would seek care for six symptoms, %	65.7*	47.6
Mean no. of myocardial infarction symptoms known (SD)	2.5 (1.3)*	3.0 (1.4)
Help-seeking		
Sought care for chest pain,[a] %	69.5	70.5
Sought care for SOB,[a] %	63.8	59.0
Median delay in seeking care for chest pain,[b] h (SD)	49.0 (388 565)	73.0 (176 799)
Median delay in seeking care for SOB,[b] h (SD)	96.0 (176 964)**	336.0 (573 968

Note. SOB = shortness of breath.
[a]Respondents with this symptom only.
[b]Help-seeking respondents only.
*Racial difference significant, $P \leq .001$.
**Racial difference significant, $P \leq .05$.
***Racial difference significant, $P \leq .01$.

Help-Seeking in Response to Coronary Heart Disease Symptoms

There were no racial differences in rates of help-seeking for chest pain or shortness of breath for either sex. The unadjusted odds ratio for Black men vs White men was 0.69 (95% confidence interval [CI] = 0.42, 1.14) for chest pain subjects and 1.08 (95% CI = 0.64, 1.81) for shortness of breath subjects. The corresponding odds ratios for women were 1.12 (95% CI = 0.74, 1.72) and 1.28 (95% CI = 0.82, 2.00). Adjusted racial

TABLE 2—Predictors of Help-Seeking in Response to Symptoms (Logistic Regression) and Delay by Help-Seekers (Linear Regression)

	Help-Seeking				Log of Hours Delayed			
	Chest Pain (n = 754)		Shortness of Breath (n = 585)		Chest Pain (n = 468)		Shortness of Breath (n = 303)	
Predictor	OR	95% CI	OR	95% CI	Coefficient	95% CI	Coefficient	95% CI
Demographics								
Black race								
Men only	0.80	0.46, 1.40	1.13	0.62, 2.03	0.80	−0.22, 1.82	−0.95	−2.36, 0.46
Women only	1.06	0.66, 1.71	1.08	0.65, 1.81	−0.57	−1.41, 0.27	−1.62	−2.74, −0.50
Female sex								
Whites only	1.14	0.65, 1.98	1.29	0.72, 2.33	0.65	−0.35, 1.65	−0.22	−1.59, 1.15
Blacks only	1.50	0.96, 2.36	1.24	0.76, 2.04	−0.70	−1.54, 0.14	−0.90	−2.06, 0.26
Socioeconomic status								
Currently employed	...	...	...	...	0.97	0.35, 1.59	...	...
Very difficult paying for basics	...	...	1.67	1.12, 2.48	...	...	...	...
Insurance coverage: uninsured	0.43	0.25, 0.64	...	...	...	...	...	...
Risk factors								
Current smoking	...	...	0.47	0.32, 0.68	...	...	...	...
Hypertension	1.71	1.22, 2.40	...	...	...	...	...	...
Elevated cholesterol	...	...	...	...	0.86	0.21, 1.50	...	...
Symptoms								
Serious chest pain	4.51	3.20, 6.36	...	...	−1.72	−2.39, −1.05	−0.90	−1.78, −0.02
Serious shortness of breath	...	...	3.34	2.30, 4.85	...	...	−1.92	−2.82, −1.02
Access/satisfaction								
Very difficult to reach care	...	...	...	...	...	...	−1.95	−3.71, −0.19
Very satisfied with care	...	...	1.67	1.07, 2.59	...	...	...	...
Propensity/knowledge								
Would seek care for six symptoms	...	...	1.56	1.08, 2.26	−0.95	−1.60, −0.30	−1.00	−1.88, −0.12
Myocardial infarction knowledge	1.24	1.09, 1.41	1.29	1.12, 1.49	...	...	...	...

Note. OR = odds ratio; CI = confidence interval.

differences (Table 2) tended to be even smaller. Results in terms of relative risk[40] (not shown here) were similar. Help-seeking was associated with insurance coverage but also with greater difficulty in paying for basics. Symptom severity and hypertension also were positively related to help-seeking, as were satisfaction with care and greater myocardial infarction knowledge. Smokers were less likely to seek care for shortness of breath. No other factors were predictive of help-seeking after these variables were controlled. Model fits were satisfactory.

Among those who sought care, Black women had a lower average (log) delay time than White women. Unadjusted effects were −0.86 (95% CI = −1.74, 0.02) and −1.84 (−3.00, −0.67) for chest pain and shortness of breath, respectively. Corresponding unadjusted effects in men were not significantly different from 0 at 0.59 (95% CI = −0.48, 1.65) and −1.14 (95% CI = −2.61, 0.34). Adjusted racial differences (Table 2) were similar in magnitude. Delay time was negatively associated with severity of symptoms and general propensity to seek care. Subjects with elevated cholesterol and current employment had longer delays in seeking care for chest pain. Subjects reporting great difficulty in reaching care had shorter delay times in seeking care for shortness of breath. Model fits were satisfactory.

Care Received by Those Who Sought Medical Attention

Among those who sought care, the relationship between race and care received differed by sex (Table 3). White men tended to have the highest rates of care and White women the lowest, particularly for recommended invasive procedures and cardiac medication, whereas rates for Black men and Black women tended to be similar. An exception was referral to a cardiologist, for which rates were significantly lower for Blacks than for Whites regardless of sex. Controlling for factors other than race did not substantially alter these patterns. The only statistically significant adjusted racial difference occurred for referral to a cardiologist (Table 4), for which the adjusted odds of referral for Blacks were roughly half the corresponding odds for Whites, regardless of sex. For recommended diagnostic procedures (Table 5) and treatments (Table 6), adjusted rates tended to be somewhat higher for Whites among men and lower for Whites among women. Results in terms of relative risk were consistent.

Adjusted differences by sex were more pronounced in Whites than in Blacks, particularly in rates of recommended treadmill tests, recommended invasive procedures, and cardiac medication. Other demographic factors related to care included age, which was positively associated with cardiac medication and diagnosis of a heart problem. Those living alone had a higher rate of hospitalization for myocardial infarction. Symptom severity and risk factors, particularly hypertension, elevated cholesterol, and diabetes, were consistently positively associated with care received, with the exceptions of current smoking and increased body mass

index. Except for current employment, which was negatively related to receipt of care, SES and insurance coverage had little consistent effect on receipt of care after other factors were controlled. Subjects reporting no difficulty in paying for basic necessities had higher rates of recommended invasive procedures. Completion of high school was negatively associated with hospitalization for myocardial infarction, and being uninsured was negatively associated with recommended electrocardiogram. Measures of access, satisfaction, and knowledge also played a small role in predicting care received among help-seekers. Having a usual source of care other than a hospital emergency room was positively associated with recommended treadmill tests but negatively associated with recommended invasive procedures. Adequate time with the doctor or nurse was positively related to being told of a heart problem. Higher levels of myocardial infarction knowledge were linked with higher rates of care. Model fits were satisfactory.

TABLE 3—Rates (%) of Coronary Heart Disease Care Received by Subjects Who Sought Medical Care, by Race and Sex

Type of Care	Black Men (n = 136)	White Men (n = 120)	Black Women (n = 305)	White Women (n = 142)
Recommended diagnostic procedures				
Electrocardiogram	88.0	86.7	91.4	92.2
Treadmill test	43.0	50.8	42.8	38.7
Echocardiogram	29.7	28.4	32.1	25.7
Treatments				
Recommended invasive procedures	20.3*	31.1	16.9	14.1
Medication	19.1	28.3	17.7	16.9
Hospitalization for myocardial infarction	18.4	22.5	9.8	13.4
Other contacts				
Referral to a cardiologist	26.9*	40.0	19.3*	29.8
Received diagnosis of heart problem	42.9	43.3	42.6	41.6

*Within-sex racial difference significant, $P < .05$.

TABLE 4—Predictors of Care Received by Help-Seeking Respondents: Other Contacts with Health Care System

Predictor	Referral to Cardiologist (n = 686)		Received Diagnosis of Heart Problem (n = 677)	
	AOR	95% CI	AOR	95% CI
Demographics				
Black race				
Men only	0.56	0.31, 0.99	1.12	0.65, 1.96
Women only	0.54	0.33, 0.91	1.12	0.71, 1.78
Female sex				
Whites only	0.56	0.32, 0.98	0.90	0.52, 1.54
Blacks only	0.54	0.32, 0.91	0.90	0.57, 1.42
Age[a]	...	...	1.22	1.09, 1.36
Socioeconomic status: currently employed	0.49	0.33, 0.72	0.65	0.44, 0.94
Risk factors				
Diabetes	...	...	1.87	1.22, 2.55
Hypertension	1.67	1.10, 2.53	1.67	1.17, 2.37
Elevated cholesterol	1.99	1.36, 2.93	...	...
Symptoms				
No chest pain	0.36	0.21, 0.64	0.33	0.21, 0.54
No shortness of breath	0.46	0.30, 0.71	0.65	0.45, 0.94
Satisfaction with care: time with physician/nurse	...	...	2.50	1.16, 5.38
Propensity/knowledge: myocardial infarction knowledge	1.26	1.10,1.43	1.15	1.02,1.30

Note. AOR = adjusted odds ratio; CI = confidence interval.
[a]Effect of 5-year increase.

Discussion

We found that, in an urban low-SES population of Blacks and Whites who were similar with respect to SES and access to care, there were few racial differences in coronary heart disease–related care patterns, either in help-seeking behavior or in care received. Although racial differences varied by sex, rates of help-seeking and care received generally did not differ significantly for Blacks and Whites. Exceptions were delay in seeking care (delays were somewhat shorter for Blacks, particularly for women) and referral to a cardiologist (rates were lower for Blacks). These racial differences persisted after other factors were controlled. The general lack of racial differences in care patterns even before adjustment for other factors may have resulted from the similarity of Black and White subjects with regard to important predictors such as severity of symptoms and employment status.

Although help-seeking behavior was similar for symptomatic men and women, help-seeking women received less care, particularly among Whites. As might be expected, the strongest and most consistent predictors of health care outcomes were risk factors and perceived seriousness of symptoms. The positive relationship between difficulty in paying for basic necessities and help-seeking was consistent with Aday and Andersen's[5] finding that low-income subjects had more hospital admissions and physician visits than those with high incomes, partly because of poorer health. Respondents who were currently employed, and hence likely to be younger or healthier than others, also had longer delays. Shorter delays among those

TABLE 5—Predictors of Care Received by Help-Seeking Respondents: Recommended Diagnostic Procedures

Predictor	Recommended Electrocardiogram (n = 687)		Recommended Treadmill Test (n = 692)		Recommended Echocardiogram (n = 668)	
	AOR	95% CI	AOR	95% CI	AOR	95% CI
Demographics						
Black race						
Men only	1.59	0.71, 3.53	0.74	0.43, 1.26	1.08	0.59, 1.97
Women only	1.03	0.48, 2.23	1.16	0.75, 1.79	1.30	0.80, 2.14
Female sex						
Whites only	1.78	0.76, 4.17	0.57	0.34, 0.97	0.88	0.49, 1.61
Blacks only	1.59	0.57, 2.35	0.90	0.58, 1.38	1.07	0.66, 1.74
Socioeconomic status: currently employed	0.40	0.23, 0.70	0.68	0.49, 0.94	0.53	0.37, 0.77
Insurance coverage: uninsured	0.43	0.20, 0.94	...	...	...	...
Risk factors						
Current smoking	0.45	0.26, 0.75	0.57	0.40, 0.80	...	...
Hypertension	...	...	...	...	1.64	1.11, 2.43
Elevated cholesterol	2.08	1.09, 3.97	2.11	1.51, 2.94	2.26	1.57, 3.26
Symptoms						
No chest pain	...	...	0.45	0.29, 0.70	0.34	0.20, 0.58
No shortness of breath	...	...	0.43	0.30, 0.61	0.43	0.29, 0.65
Access to care: usual source of care	...	...	1.86	1.11, 3.11	...	...
Propensity/knowledge: myocardial infarction knowledge	1.44	1.16, 1.80	...	...	...	...

Note. AOR = adjusted odds ratio; CI = confidence interval.

reporting great difficulty in reaching care may reflect the need to compensate for longer times to reach care. The lack of association between help-seeking and measures of access and satisfaction may have been due in part to respondents' seeking care from sources other than their usual source (e.g., hospital emergency rooms). Negative associations between care received and current employment may reflect the "healthy worker effect"; that is, subjects who received care may have been less likely to be healthy enough to continue working than were subjects who did not receive care. Similarly, those who received care may have been instructed to stop smoking or to lose weight and hence were less likely to continue these behaviors than were those who did not receive care.

Adjustment for socioeconomic factors and insurance coverage had little impact on Black–White differences in care, possibly because racial differences in SES in our sample were somewhat smaller than in the general population: 1980 median census tract annual household income averaged approximately $14 000 for Whites and $11 000 for Blacks in our sample, compared with average annual household income of approximately $19 000 for Whites and $11 000 for Blacks nationally.[41] Although racial differences in SES in this sample were statistically significant, the gap may not have been large enough to yield SES-related racial differences in care patterns, particularly given the similarities of Blacks and Whites with respect to other predictors. In more heterogeneous study populations, SES and insurance coverage may account for more racial differences in coronary heart disease care than we found in our sample. The longer delay times for Blacks than for Whites found by Cooper and colleagues,[4] for example, may have resulted in part from the comparison of poor and working-class Blacks admitted to an inner-city hospital with Whites in more affluent communities. Oberman and Cutter proposed racial differences in education, income, and occupation as possible explanations for lower rates of coronary artery bypass grafting among Blacks.[10] Much of the racial difference in rates of surgery found in the Coronary Artery Surgery Study was due to Whites for whom medical therapy was recommended but who chose surgical treatment,[13] perhaps reflecting Black–White differences in ability to pay for care.

Previous work on help-seeking found lower rates of help-seeking for chest pain[3] and higher rates of delay among nonobstetric and nonpsychiatric hospital admissions[29] in Blacks than in Whites. Neither study, however, adjusted these racial differences for perceived severity of symptoms. Consistent with past research,[9–13] we found lower adjusted rates of recommended invasive procedures in Black men than in White men, although the difference was not statistically significant ($P = .15$). As in recent studies of care patterns by sex,[42,43] the women in our sample, particularly White women, were less likely than men to have recommended invasive procedures even after other relevant factors were controlled. Recommended invasive procedures were positively associated with reporting no difficulty in paying for basics, similar to previous findings that privately insured patients had higher rates of cardiac surgery.[44] In addition, Blacks were significantly less likely than Whites to be referred to a cardiologist, consistent with past findings that Blacks had lower rates of referral to specialists.[45]

Our findings must be interpreted with caution. First, outcome and predictor measures were self-reported and hence

TABLE 6—Predictors of Care Received by Help-Seeking Respondents: Treatment

Predictor	Recommended Invasive Procedures (n = 671)		Cardiac Medication (n = 664)		Hospitalized for Myocardial Infarction (n = 678)	
	AOR	95% CI	AOR	95% CI	AOR	95% CI
Demographics						
Black race						
Men only	0.61	0.31, 1.19	0.81	0.43, 1.54	1.02	0.49, 2.11
Women only	1.27	0.67, 2.42	1.36	0.78, 2.39	0.94	0.46, 1.92
Female sex						
Whites only	0.32	0.16, 0.64	0.54	0.29, 1.01	0.44	0.21, 0.92
Blacks only	0.67	0.37, 1.20	0.91	0.52, 1.59	0.40	0.21, 0.78
Age[a]	...	...	1.33	1.18, 1.50	1.36	1.17, 1.58
Living alone	...	...	...	...	2.08	1.24, 3.46
Socioeconomic status						
Completed high school	...	...	...	...	0.57	0.34, 0.95
Currently employed	0.44	0.28, 0.69	...	...	...	...
No difficulty paying for basics	1.63	1.01, 2.62	...	...	...	...
Risk factors						
Current smoking	0.59	0.37, 0.96	...	...	...	...
Stopped smoking	...	...	...	...	2.31	1.31, 4.07
Diabetes	2.33	1.39, 3.91	...	...	...	...
Family history	...	...	1.80	1.21, 2.70	...	...
Log body mass index[b]	0.59	0.48, 0.72	...	...	...	...
Hypertension	1.95	1.19, 3.21	...	...	...	...
Elevated cholesterol	2.12	1.37, 3.27	...	...	2.13	1.32, 3.44
Symptoms						
No chest pain	0.14	0.06, 0.33	0.40	0.22, 0.72	0.10	0.03, 0.30
No shortness of breath	0.48	0.29, 0.78	0.49	0.31, 0.76	0.40	0.23, 0.68
Access to care: usual source of care	0.52	0.28, 0.95	...	...	...	...
Propensity/knowledge: myocardial infarction knowledge	...	...	...	...	1.29	1.08, 1.52

Note. AOR = adjusted odds ratio; CI = confidence interval.
[a]Effect of 5-year increase.
[b]Effect of increase from 25th percentile (3.18) to 75th percentile (3.46).

subject to problems inherent in a community-based sample, such as recall bias and misunderstanding of survey questions. Moreover, as Strogatz[3] points out, perception and report of chest pain and other symptoms may differ for Blacks and Whites and thus may not have been adequately addressed by the survey questions regarding symptoms. Any such racial differences in reporting of symptoms could have affected the estimation of racial differences in care patterns. Self-reports of medical conditions, however, including reports of chronic heart conditions,[46–49] have been found to have a fairly high level of agreement with medical records. In addition, the self-reported data collected here did not include clinical measures, which would inform clinical decision making. We note, however, that no judgments were made in this study regarding appropriateness of care received.

Second, it was not feasible to ascertain the timing of outcomes relative to timing of occurrence of symptoms and other predictors. As noted previously,[3] perceptions or reports of symptoms and risk factor status, as well as measures of access, satisfaction, and knowledge, may be results rather than causes of seeking and/or receiving care. When the models for help-seeking and receipt of care were reestimated with measures of access, satisfaction, and knowledge omitted, however, adjusted racial differences changed very little. A related problem is that predictors such as insurance coverage were current measures and hence may not have been relevant to care patterns for symptoms first occurring years earlier. Analyses restricted to subjects with symptom onset less than a year before the interview, however, yielded similar results.

The results presented here raise two main issues needing further examination. First, it is not clear why help-seeking Blacks were significantly less likely than help-seeking Whites to have been referred to a cardiologist, even after risk factors, severity of symptoms, and ability to pay for care were controlled. This result is particularly surprising considering the lack of Black–White differences in reported treatments and procedures normally associated with cardiology practice. When a low significance level (.01) was substituted for a level of .05 to compensate for multiple hypothesis testing,[50] this racial difference remained statistically significant. Second, White women had consistently lower rates of receipt of care than the other three race–sex groups, and women of both races had lower rates of referral to a cardiologist than did men, even after other factors were accounted for. Further study is needed to determine whether these differences can be explained by factors not adequately measured in this study, such as clinical data, or whether they reflect true racial and sex differences in receipt of coronary heart disease–related care. □

Acknowledgments

This work was supported by grant HL39232 from the National Heart, Lung, and Blood Institute.

References

1. Gillum RF. Coronary heart disease in black populations, I: mortality and morbidity. *Am Heart J.* 1982;104:839–851.
2. Maynard C, Fisher LD, Passamani ER, Pullum T. Blacks in the Coronary Artery Surgery Study (CASS): risk factors and coronary artery disease. *Circulation.* 1986; 74:64–71.
3. Strogatz DS. Use of medical care for chest pain: differences between Blacks and Whites. *Am J Public Health.* 1990;80:290–294.
4. Cooper RS, Simmons B, Castaner A, Prasad R, Franklin C, Ferlinz J. Survival rates and prehospital delay during myocardial infarction among black persons. *Am J Cardiol.* 1986;57:208–211.
5. Aday LA, Andersen RM. The national profile of access to medical care: where do we stand? *Am J Public Health.* 1984;74: 1331–1339.
6. Davis K, Gold M, Makue D. Access to health care for the poor: does the gap remain? *Ann Rev Public Health.* 1983;2:159–182.
7. James SA, Wagner EH, Strogatz DS, et al. The Edgecombe County high blood pressure control program, II: barriers to the use of medical care among hypertensives. *Am J Public Health.* 1984;74:468–472.
8. *Office Visits for Diseases of the Circulatory System: The National Ambulatory Medical Care Survey, United States, 1975–1976.* Hyattsville, Md: National Center for Health Statistics; 1979. DHHS publication PHS 79-1791.
9. Graves E, Haupt BJ. *Detailed Diagnoses and Surgical Procedures for Patients Discharged from Short-Stay Hospitals, United States, 1977.* Washington, DC: US Dept of Health and Human Services; January 1979. DHHS publication PHS 79-1274.
10. Oberman A, Cutter G. Issues in the natural history and treatment of coronary heart disease in black populations: surgical treatment. *Am Heart J.* 1984;108:688–694.
11. Ford E, Cooper R, Castaner A, et al. Coronary arteriography and coronary bypass surgery among whites and other racial groups relative to hospital-based incidence rates for coronary artery disease: findings from NHDS. *Am J Public Health.* 1989;79: 437–440.
12. Wenneker MB, Epstein AM. Racial inequalities in the use of procedures for patients with ischemic heart disease in Massachusetts. *JAMA.* 1989;261:253–257.
13. Maynard C, Fisher LD, Passamani ER, Pullum T. Blacks in the Coronary Artery Surgery Study: race and clinical decision making. *Am J Public Health.* 1986;76:1446–1448.
14. Gillum RF. Coronary artery bypass surgery and coronary angiography in the United States, 1979–1983. *Am Heart J.* 1987;113: 1255–1260.
15. Bombardier C, Fuchs VR, Lillard LE, et al. Socioeconomic factors affecting the utilization of surgical operations. *N Engl J Med.* 1977;297:699–705.
16. Young LY. Who is likely to have coronary bypass surgery? Presented at the annual meeting of the American Statistical Association; August 15–18, 1983; Toronto, Ont, Canada.
17. Adams L, Africano E, Doswell W, et al. Summary of Workshop I: Working Group on Epidemiology. *Am Heart J.* 1984;108: 699–702.
18. Keil JE, Loadholt CB, Weinrich MC, Sandifer SH, Boyle E. Incidence of coronary heart disease in blacks in Charleston, South Carolina. *Am Heart J.* 1984;108:779–786.
19. Lee MH, Namat OB, Kuller LH. Validation of reported myocardial infarction mortality in blacks and whites: a report from the Community Cardiovascular Surveillance Program. *Ann Epidemiol.* 1990;1: 1–12.
20. Maynard C, Fisher LD, Passamani ER. Survival of black persons compared with white persons in the Coronary Artery Surgery Study (CASS). *Am J Cardiol.* 1987;60:513–518.
21. Simmons BE, Castaner A, Mar M, Islam N, Cooper R. Survival determinants in black patients with angiographically defined coronary artery disease. *Am Heart J.* 1990;119:513–519.
22. Haywood LJ. Coronary heart disease mortality/morbidity and risk in blacks, I: clinical manifestations and diagnostic criteria: the experience with the Beta Blocker Heart Attack Trial. *Am Heart J.* 1984;108: 787–793.
23. Tofler GH, Stone PH, Facc JE, et al. Effects of gender and race on prognosis after myocardial infarction: adverse prognosis for women, particularly black women. *J Am Coll Cardiol.* 1987;9:473–482.
24. Sempos C, Cooper R, Kovar MG, McMillen M. Divergence of the recent trends in coronary mortality for the four major race-sex groups in the United States. *Am J Public Health.* 1988;78:1422–1427.
25. McCord C, Freeman HP. Excess mortality in Harlem. *N Engl J Med.* 1990;322:173–177.
26. Levy M. Black/white comparisons of premature mortality for public health program planning—District of Columbia. *MMWR.* 1989;38:33–37.
27. Keil JE, Saunders DE, Lackland DT, et al. Acute myocardial infarction: period prevalence, case fatality, and comparison of black and white cases in urban and rural areas of South Carolina. *Am Heart J.* 1985;109:776–784.
28. Feinleib M, Davidson MJ. Coronary heart disease mortality: a community perspective. *JAMA.* 1972;222:1129–1134.
29. Weissman JS, Stern R, Fielding SL, Epstein AM. Delayed access to health care: risk factors, reasons, and consequences. *Ann Intern Med.* 1991;114:325–331.
30. Mutchler JE, Burr JA. Racial differences in health and health care service utilization in later life: the effect of socioeconomic status. *J Health Soc Behav.* 1991;32:342–356.
31. Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. *Med Care.* 1981;19: 127–140.
32. Koos E. *The Health of Regionsville: What the People Thought and Did About It.* New York, NY: Columbia University Press; 1954.
33. Draper N, Smith H. *Applied Regression Analysis.* 2nd ed. New York, NY: John Wiley & Sons Inc; 1981.
34. Glueck CJ, Gartside P, Laskarzewski PM, Khoury P, Tyroler HA. High-density lipoprotein cholesterol in blacks and whites: potential ramifications for coronary artery disease. *Am Heart J.* 1984;108:815–826.
35. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons Inc; 1989.
36. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health.* 1991;81:1630–1635.
37. *SAS Technical Report P-200, SAS/STAT Software: CALIS and LOGISTIC Procedures, 6.04.* Cary, NC: SAS Institute Inc; 1990.
38. *BMDP Statistical Software Manual.* Berkeley, Calif: University of California Press; 1988.
39. Kish L. *Survey Sampling.* New York, NY: John Wiley & Sons Inc; 1965.
40. Rosner B. *Fundamentals of Biostatistics.* 3rd ed. Boston, Mass: PWS-Kent Publishing Co; 1989.
41. *Statistical Abstract of the United States: 1990.* Washington, DC: US Dept of Commerce, Bureau of the Census; 1990.
42. Ayanian JZ, Epstein AM. Differences in the use of procedures between women and men hospitalized for coronary heart disease. *N Engl J Med.* 1991;325:221–225.
43. Steingart RM, Packer M, Hamm P, et al. Sex differences in the management of coronary artery disease. *N Engl J Med.* 1991;325:226–230.
44. Wenneker MB, Weissman JS, Epstein AM. The association of payer with utilization of cardiac procedures in Massachusetts. *JAMA.* 1990;264:1255–1260.
45. Penchansky R, Fox D. Frequency of referral and patient characteristics in group practice. *Med Care.* 1970;8:368–385.
46. Bush TL, Miller SR, Golden AL, Hale WE. Self-report and medical record report agreement of selected medical conditions in the elderly. *Am J Public Health.* 1989;79: 1554–1556.
47. Fienberg SE, Loftus EF, Tanur JM. Recalling pain and other symptoms. *Milbank Memorial Fund Q Health Soc.* 1985;63:582–597.
48. O'Toole BI, Battistutta D, Long A, Crouch K. A comparison of costs and data quality of three health survey methods: mail, telephone and personal home interview. *Am J Epidemiol.* 1986;124:317–328.
49. Weeks MF, Kulka RA, Lessler JT, Whitmore RA. Personal versus telephone surveys for collecting household health data at the local level. *Am J Public Health.* 1983;73: 1389–1394.
50. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22:719–748.

Acute Chest Pain in African Americans: Factors in the Delay in Seeking Emergency Care

Kathleen Ell, DSW, L. Julian Haywood, MD, Eugene Sobel, PhD, Maria deGuzman, MD, David Blumfield, MD, and Jia-Ping Ning, BA

ABSTRACT

Objectives. African Americans have been shown to have longer delay times than the majority population in seeking care for acute cardiac problems. The purpose of this study was to determine whether socioeconomic factors affect delay times.

Methods. Structured interviews were administered to 254 African Americans admitted to a public hospital and 194 African Americans admitted to a private hospital for suspected acute myocardial infarction.

Results. Patient characteristics found by multiple regression analysis to affect decision-making and travel time for care-seeking were structural access to care, persistence of symptoms, degree of incapacitation, consultation with a layperson, consultation with medical professionals, and mode of transportation.

Conclusions. Within-group differences were found to be related to socioeconomic status. Strategies to increase knowledge about heart attack symptoms, improve access to care, and improve the socioeconomic status of at-risk African Americans are indicated. (*Am J Public Health.* 1994;84:965–970)

Introduction

According to many studies, African Americans have higher rates of morbidity and mortality from ischemic heart disease than US Whites, both as inpatients and as outpatients.[1–11] Treatment delay blunts the effectiveness of new therapies,[12,13] and African Americans have been shown to receive less specialized care, including coronary artery bypass graft surgery.[14–17] Inordinately long delay times have been documented in this population group,[18] and socioeconomic status (SES) may be an added factor.[19–22] Symptom response is said to be influenced by social and psychological factors including demographic characteristics,[23] previous experience with the medical care system,[24] perceptions of symptom quality and severity,[25] beliefs about care-seeking appropriateness[23] and effectiveness,[24] and accessibility of care. Whether African-American patients who seek prompt care have different characteristics than those who do not is not clear.

We report results from a large study of health care–seeking behavior for acute chest pain. We compare the duration of care-seeking behavior in African Americans of lower and middle SES. We examine associations between demographic and medical care utilization characteristics and the duration of care-seeking behavior among African-American patients who sought care in an emergency room and were subsequently hospitalized for acute chest pain. Duration of care-seeking refers to the length of time between symptom onset and arrival at the hospital. Consistent with the perspective that patients engage in social and psychological processes in deciding to seek emergency medical care,[25] the duration between symptom onset and the decision to seek emergency care (decision phase) is examined separately from the time between decision to seek care and hospital arrival (travel logistics phase).

Methods

Study Sample

Between August 1988 and July 1990, three consecutive case series of African-American, Latino, and non-Latino White patients hospitalized for acute chest pain were recruited from two Los Angeles medical centers: a large urban public hospital where the majority of patients had no insurance and often no other source of medical care, and a large urban private health maintenance organization (HMO) hospital where virtually all patients had insurance and a regular source of care. In this report we examine the African-American sample, which consisted of 254 patients from the public hospital and 194 patients from the private hospital. Because they were recruited consecutively, these patients are presumed to be representative of those receiving care at each medical center. Exclusionary criteria for patients of all three racial/ethnic groups were the following: unwillingness to give consent, severity

Kathleen Ell is with the School of Social Work, University of Southern California, Los Angeles. L. Julian Haywood, Eugene Sobel, Maria deGuzman, and Jia-Ping Ning are with the School of Medicine, University of Southern California, Los Angeles. David Blumfield is with Kaiser Permanente Hospital, Los Angeles.

Requests for reprints should be sent to L. Julian Haywood, MD, LAC+USC Medical Center, 1200 N State St, Box 305, Los Angeles, CA 90033.

This paper was accepted November 3, 1993.

TABLE 1—Demographic and Medical Care Utilization Characteristics (%) of African-American Patients, by Type of Hospital

	Public Hospital (n = 254)	Private (n = 194)	P
Sex			
Male	52	51	NS
Female	48	49	
Marital status			
Married/living together	23	58	.0001
Unmarried	77	42	
Age, y			
< 44	26	13	.0001
45–64	56	50	
≥ 65	18	37	
Education, y			
≤ 8	19	10	.0001
9–12	55	37	
≥ 13	26	53	
Hollingshead Index[a]			
1–3	25	59	.0001
4–5	75	41	
Employment			
Full-time	18	47	.0001
Part-time	17	4	
Unemployed	5	1	
Retired	11	32	
Disabled	42	12	
Homemaker	6	4	
Monthly income			
< $1000	97	19	.0001
> $1000	3	81	
Health insurance			
Yes	25	99	.0001
No	75	1	
Medical care utilization			
Regular source of care	72	96	.0001
Waiting time ≥ 1 h	62	0	.0001
Physician spends enough time	86	92	.04
Regular physician visits	58	54	NS
Previous emergency room visits	46	31	.001

[a]Hollingshead Two-Factor Index of Social Position.[26]

of illness, inability to cooperate because of mental status, and alternative clinical diagnoses (i.e., acute pneumonia with pleuritic chest pain, acute costrochondritis). The most common reason for nonparticipation of eligible patients was discharge prior to being interviewed. Participation rates were 58% for African-American patients screened at the public hospital and 65% for those at the private hospital.

Data Collection

All patients were interviewed in the hospital by a bilingual, bicultural interviewer. A detailed structured questionnaire incorporating instruments previously used in surveys of health care behavior and access to care was completed for each study subject. In this report we analyze data from the questionnaire that pertain to the lengths of the decision and travel logistics phases: the former includes time spent in interpreting symptoms, trying self-care methods, and consulting laypersons or medical professionals, and the latter includes time spent in obtaining transportation and actually traveling to the hospital. To record the duration of the care-seeking phases, interviewers verified times by using a benchmark technique whereby the patient was asked to verify the times reported with an event occurrence, daily routine, or break in routine.

Patient characteristics previously shown or hypothesized to influence the duration of care-seeking behavior for acute symptoms were also examined. These included sex, age, marital status, previous history of coronary heart disease, perceived general health status, experience of a prodromal episode prior to the time of acute onset, having regularly seen a doctor within the past year, having had an emergency room visit within the past 6 months, and patients' reported perception and causal attribution of the acute symptoms. These data were obtained with an existing instrument.[25] SES indicators include education and the Hollingshead Two-Factor Index of Social Position.[26] Measures of access to care include health insurance, a regular source of care, and hospital site.

Data Analysis

Bivariate associations were examined with chi-square Wilcoxon and Kruskal-Wallis rank sum statistics and two-way analysis of variance to assess relationships between duration phases and other variables. Multiple regression modeling methods (including both forward inclusion and backward elimination) were used to compare results between study cohorts.[27,28] In view of the wide ranges and skewed distributions of duration measurements, the median test is reported along with mean difference in bivariate analyses; log transformations for the duration variables were used in the multivariate analyses. If care-seeking duration exceeded 1 week, or if the subject reported travel time to the hospital as 0 hours, the subject was deleted from the analyses (n = 9 public hospital and 3 private hospital patients).

Results

The sampling design was expected to yield significantly different patient populations with respect to SES and access to care as well as overall experiences with medical care and previous encounters with health care systems. The data presented in Table 1 show that compared with African-American patients in the private hospital, those in the public hospital were disadvantaged with respect to all measured SES indicators (income, Hollingshead Index, and education), less likely to have health insurance or a regular source of care, and more likely to have had a previous emergency room visit. The public-hospital patients also reported less satisfaction with previous encounters with the health care system. These results support the hypothesis that hospital site is an indicator of access to care.

The total study population mean, median, and range for the time between onset of symptoms and the decision to seek emergency care were 9.14 hours, 1.90 hours, and 0 to 122 hours, respectively; corresponding values for the time to arrive at the hospital following the decision to seek care (i.e., the travel logistics phase) were 5.74 hours, 0.75 hours, and 0.2 to 169 hours. The mean and median durations of the decision and travel logistics phases were compared by SES and access-to-care indicators, sex, previous health status, characteristics of the acute symptoms, and patient behaviors (Table 2). The results demonstrate significant within-group differences among African-American patients. Public hospital patients, those without health insurance, and those who reported usual waits of 1 hour or more to see a physician reported longer times in reaching the decision to seek care. Women experienced a significantly longer travel logistics phase and a longer, but not significantly longer, decision phase than men. Care-seeking durations did not vary between patients having a regular source of care and those not having a regular source of care. The results indicate a trend toward longer times before seeking care among the less educated and lower-SES patients, but these differences were not significant.

Previous history of angina, acute myocardial infarction, or experience of a prodromal symptom phase did not result in significant differences in care-seeking duration; however, care-seeking times were shorter for those with a positive history. In contrast, self-reported health status and being alone at the time of acute symptom onset were associated with a longer decision-phase time. Whether patients attributed their symptoms to a particular cause or did not know the possible cause did not affect the decision-phase duration but appeared to influence the duration of the travel logistics phase. However, when patients perceived their chest pain to be continuous and incapacitating, the duration of care-seeking was significantly shorter than when patients perceived the pain to be intermittent or nonincapacitating (Table 2).

The following specific patient behaviors also influenced care-seeking duration: engaging in self-care by taking medications, informing other members of one's social network about the symptoms, calling a physician or paramedics for consultation, and mode of transportation.

To examine the relative contribution of factors found to distinguish care-seeking duration, multiple regression analyses were conducted. With the variables found to be significant in Table 2 (the variable "waiting time more than 1 hour" was excluded from this analysis because it was not applicable to patients without a regular source of care), the results of analyses using forward inclusion and backward elimination techniques were identical (Table 3). Hospital site, severity of acute symptoms, informing a layperson, consulting a medical professional, and mode of transportation were significant predictors of the duration of the decision phase; having informed a layperson and transportation mode were significant predictors of the duration of the travel logistics phase. Separate analyses for each hospital site (not shown) found that being alone, degree of incapacitation, and transportation were significant predictors of decision-phase duration among public hospital patients ($R^2 = .24$); having told a layperson, symptom pattern, degree of incapacitation, and having taken medication were significant predictors of decision-phase duration among private patients ($R^2 = .18$). Mode of transportation predicted the travel time for each site.

Discussion

This study is the first reported detailed examination of African-American care-seeking behavior for acute chest pain in which comparisons are made between lower- and middle-SES patients and between patients with different levels of structural access to medical care. These results enhance our knowledge about the duration of prehospital behavior, including time spent in deciding to seek medical care and time spent traveling to the hospital, among African Americans who decide to seek emergency care and are subsequently hospitalized. Our findings for African-American patients are similar to other researchers' findings for predominantly White populations,[13,25] but we also uncovered differences specific to the care-seeking behavior of African Americans by access to care, SES, and sex.

The total prehospital median duration found in this study of 2.65 hours (decision phase time = 1.90 hours, travel logistics phase time = 0.75 hours) is longer than the 2.00 hours reported in a study conducted in a predominantly White population.[13] In contrast, an earlier study of African-American patients at Cook County Hospital reported a mean prehospital duration of 20.2 hours, compared with 16.22 for public hospital patients in this study, and a median of 6.4 hours for hospitalized patients with acute myocardial infarction.[18] It is noteworthy that in our study, the mean and median durations of the decision phase were significantly longer for patients receiving care at the public hospital than for those receiving care at the private hospital. We also found that patients without health insurance had a significantly longer decision phase time than those with health insurance. These results are consistent with results from a survey of 12 068 patients hospitalized for various reasons in which it was found that delay in obtaining medical care was 40% to 80% greater for patients who were African-American, poor, uninsured, or without a regular physician.[29]

As found in previous studies,[13,18,25] the time patients spend in reaching the decision to seek emergency medical care accounts for the largest proportion of the total duration of care-seeking behavior. Results from the bivariate analyses shown in Table 2 also indicate a trend in which lower socioeconomic class and limited education further define African Americans who delay the decision to obtain medical care. In addition, our results indicate that African-American women experienced longer prehospital durations following symptom onset than did men. This finding is noteworthy (as well as consistent with results from other studies[18,30]) because women, particularly African-American women, are at greater risk than men for poor coronary heart disease outcomes.[4,5,8,18,29]

The results of the multivariate analyses demonstrate that structural access to care, as indicated by hospital site, was a significant predictor of duration of the decision phase. Our findings provide evidence that socioeconomic factors may influence the care-seeking behavior of African Americans, primarily through a relationship to access to care.

Patients' symptom perceptions and whether they told others in their social network and/or consulted medical professionals were also significant predictors of the duration of the decision phase. These results are consistent with previous research[25,30–32] and indicate that the decision to seek medical care by African Americans is influenced by factors similar to those found to influence the decision in predominantly White populations.[13]

Public education[33–35] about the importance of cardiac symptomatology may have helped influence the care-seeking behavior of patients in our sample. On the

TABLE 2—Duration of Care-Seeking Behavior, by Demographic, Utilization, Clinical History, Symptom, and Patient Behavior Characteristics

	Decision Phase[a]				Travel Logistics Phase[b]			
	Mean, h	95% CI	Median, h	95% CI	Mean, h	95% CI	Median, h	95% CI
Demographics								
Hospital								
Public	11.02*	8.1, 13.9	2.50*	1.8, 3.2	4.20	2.1, 6.3	0.83	0.7, 1.0
Private	6.99	4.9, 9.1	1.07	0.4, 1.7	7.50	4.1, 10.9	0.67	0.6, 0.8
Health insurance								
Yes	7.83**	5.6, 10.0	1.00**	0.6, 1.4	6.54	3.8, 9.3	0.67	0.6, 0.8
No	11.15	8.0, 14.3	2.67	1.5, 3.8	4.63	2.0, 7.3	1.00	0.9, 1.1
Hollingshead Index[c]								
1–3	5.79	4.2, 7.4	1.50	0.7, 2.3	6.35	3.2, 9.5	0.68	0.4, 0.9
4–5	11.58	8.7, 14.5	2.00	1.2, 2.8	5.30	2.9, 7.7	0.75	0.6, 0.9
Education, y								
≤8	12.26	5.5, 19.1	1.96	0.1, 3.8	4.93	−0.1, 9.9	1.00	0.8, 1.2
9-12	9.39	6.8, 12.0	2.00	1.4, 2.6	4.77	2.2, 7.3	0.75	0.6, 0.9
>12	7.68	5.2, 10.2	1.50	0.7, 2.3	7.16	3.7, 10.6	0.68	0.4, 0.9
Sex								
Male	7.12	4.9, 9.3	1.67	1.0, 2.4	5.20	2.8, 7.6	0.67*	0.6, 0.8
Female	11.36	8.4, 14.4	2.00	1.1, 2.9	6.34	3.3, 9.4	0.92	0.8, 1.1
Age, y								
<65	10.21	7.9, 12.5	2.00	1.4, 2.6	6.15	3.8, 8.5	0.75	0.7, 0.8
≥65	5.73	3.7, 7.7	1.29	0.2, 2.4	4.45	1.5, 7.4	0.79	0.6, 1.0
Medical care utilization								
Regular source of care								
Yes	8.02	6.2, 9.8	1.05	0.7, 1.4	5.36	3.4, 7.3	0.60	0.5, 0.7
No	7.83	3.9, 11.7	1.90	0.2, 3.6	3.38	1.0, 5.7	0.80	0.5, 1.1
Waiting time ≥1 h								
Yes	13.35*	8.5, 18.2	3.38**	2.1, 4.7	5.33	1.0, 9.6	1.00**	0.9, 1.1
No	7.68	5.5, 9.8	1.25	0.8, 1.7	6.67	3.9, 9.4	0.67	0.6, 0.8
Physician spends enough time								
Yes	9.46	7.2, 11.7	1.50	0.8, 2.2	6.52	4.0, 9.0	0.72	0.7, 0.8
No	7.60	3.3, 11.9	2.50	0.4, 4.6	4.45	−0.7, 9.6	1.00	0.6, 1.4
Regular physician visits								
Yes	9.97	7.2, 12.7	2.00	1.2, 2.8	7.39	4.2, 10.6	0.83	0.7, 1.0
No	8.16	5.8, 10.5	1.67	1.0, 2.4	3.81	2.0, 5.6	0.67	0.6, 0.8
Previous emergency room visit								
Yes	9.25	6.3, 12.2	1.96	1.3, 2.7	3.82	1.4, 6.3	0.71	0.6, 0.8
No	9.10	6.7, 11.5	1.83	1.0, 2.7	6.96	4.2, 9.7	0.83	0.6, 1.0
Clinical history								
Acute myocardial infarction								
Yes	7.89	4.8, 11.0	1.37	0.7, 2.1	4.47	1.7, 7.3	0.75	0.6, 0.9
No	9.90	7.5, 12.3	2.00	1.3, 2.7	5.78	3.5, 8.1	0.75	0.6, 0.9
Angina								
Yes	7.09	5.0, 9.1	1.54	0.8, 2.2	5.86	3.1, 8.6	0.76	0.6, 1.0
No	11.91	8.7, 15.1	2.00	0.7, 3.3	5.71	3.0, 8.4	0.67	0.4, 0.9
Prodromal phase								
Yes	7.92	5.8, 10.1	2.00	1.2, 2.8	5.09	2.6, 7.6	0.88*	0.7, 1.0
No	10.13	7.3, 13.0	1.67	0.9, 2.5	6.28	3.4, 9.1	0.67	0.6, 0.8
Self-reported health status								
Good	7.13*	5.2, 9.1	1.23*	0.6, 1.8	5.82	3.1, 8.6	0.65*	0.5, 0.8
Poor	11.39	8.2, 14.6	2.78	1.9, 3.6	5.66	3.0, 8.3	1.00	0.9, 1.1
Acute symptoms								
Attribution								
Yes	7.88	5.7, 10.0	2.00	1.2, 2.8	5.07*	2.6, 7.5	0.92*	0.8, 1.1
No	10.09	7.2, 13.0	1.67	0.9, 2.5	6.32	3.4, 9.2	0.67	0.6, 0.8
Pattern								
Intermittent	12.29***	8.7, 15.9	3.38**	1.5, 5.3	4.50	2.3, 6.7	0.83	0.6, 1.0
Continuous	5.23	3.7, 6.8	1.33	0.9, 1.8	6.52	3.3, 9.7	0.72	0.6, 0.8
Intensity								
Increasing, stable	6.49	4.2, 8.7	1.83	1.2, 2.5	4.63	2.5, 6.7	0.75	0.6, 0.9
Decreasing	7.14	4.6, 9.7	2.25	0.5, 4.0	7.87	3.1, 12.7	0.68	0.4, 0.9
Incapacitation								
Little, none	11.37***	8.4, 14.3	2.92***	1.8, 4.1	5.32	2.7, 7.9	0.83	0.7, 1.0
Severe	4.15	2.9, 5.4	1.00	0.6, 1.4	6.18	2.7, 9.6	0.62	0.5, 0.8

(Continued)

TABLE 2—*Continued*

	Decision Phase[a]				Travel Logistics Phase[b]			
	Mean, h	95% CI	Median, h	95% CI	Mean, h	95% CI	Median, h	95% CI
Patient behaviors								
Took medication								
Yes	12.58**	8.6, 16.6	3.44*	1.3, 5.6	9.42	3.4, 15.4	1.00*	−0.3, 2.3
No	7.34	5.0, 9.7	1.50	0.8, 2.2	5.26	2.9, 7.7	0.67	0.5, 0.8
Took medication after telling others								
Yes	9.60	6.6, 12.6	2.00	1.4, 2.6	5.14	2.5, 7.8	0.67	0.5, 0.8
No	8.98	2.6, 15.4	4.10	0.3, 7.9	4.38	−3.0, 11.7	0.50	0.3, 0.7
Told layperson								
Yes	9.36*	7.1, 11.6	2.08	1.2, 3.0	7.57*	4.6, 10.6	0.75	0.5, 1.0
No	8.82	5.7, 11.9	1.75	0.9, 2.6	3.18	1.3, 5.0	0.75	0.7, 0.8
Called physician or paramedics								
Yes	8.30*	5.3, 11.3	1.33	0.7, 2.0	6.23***	3.4, 9.0	0.52***	0.4, 0.6
No	9.80	7.5, 12.1	2.50	1.1, 3.9	5.36	2.7, 8.0	1.00	0.9, 1.1
Alone								
Yes	10.45**	7.2, 13.7	2.75*	1.6, 3.9	5.68	2.8, 8.6	0.75	0.6, 0.9
No	8.50	6.3, 10.7	1.25	0.7, 1.8	5.77	3.3, 8.3	0.75	0.6, 0.9
Transportation to hospital								
Paramedics	3.01***	0.9, 5.1	0.65***	0.4, 1.0	4.79***	1.5, 8.1	0.50***	0.5, 0.5
Other	10.79	8.2, 13.4	2.67	1.5, 3.8	5.38	2.9, 7.9	1.00	0.9, 1.1

[a]Time between symptom onset and decision to seek medical help.
[b]Time between decision to seek help and arrival at medical facility.
[c]Hollingshead Two-Factor Index of Social Position.[26]
*$P < .05$; **$P < .01$; ***$P < .001$.

other hand, we found that attributing acute chest pain to a heart attack or other hypothesized cause vs making no causal attribution did not predict the duration of decision making, consistent with previously reported results.[13] Furthermore, our findings raise questions about whether knowledge about cardiac symptomatology is similar among all African Americans. For example, patients receiving care at the public hospital and those without health insurance were more likely to take medication before telling others about their symptoms; this self-care behavior was associated with longer care-seeking durations.

Efforts to locate a family member for advice may forestall prompt action, and consultation with others could result in delaying the decision to obtain emergency care. Previous research found that the delay in obtaining care for acute myocardial infarction was longer when a spouse or others initiated the decision to seek medical care than when the decision was the patient's.[32] On the other hand, a family member could strongly encourage or initiate the decision to seek immediate aid. Similarly, consultation with medical professionals could result in speeding or delaying the decision to seek emergency care.[36] In this study, as in previous research, involving others in the decision-making process prolonged the time African Americans spent in reaching the decision to seek care and lengthened the travel logistics phase. Interestingly, being alone also lengthened care-seeking duration, perhaps because there was no one to act on the behalf of the patient. The wide range in the duration of the travel logistics phase suggests that many patients experience delays in getting to the hospital after deciding to seek care. Perhaps the process of enlisting non–health professionals in travel plans enhances this delay.

TABLE 3—Stepwise Multiple Regression Analyses of Decision[a] and Travel Logistics[b] Phases (Log-Transformed Data)

	Coefficient	SE	*P*
Decision phase			
Hospital (0 = public, 1 = private)	−1.0953	0.3329	.0011
Symptom pattern (0 = continuous, 1 = intermittent)	0.7435	0.2723	.0067
Degree of incapacitation (0 = severe, 1 = little or none)	0.9788	0.2752	.0004
Told layperson (0 = no, 1 = yes)	1.1171	0.3148	.0004
Consulted physician or paramedics (0 = no, 1 = yes)	0.7245	0.3224	.0253
Transportation (0 = paramedics, 1 = other)	1.8480	0.3621	.0001
Alone (0 = no, 1 = yes)	0.5410	0.2888	.0620
$R^2 = .2101$			
Travel logistics phase			
Told layperson (0 = no, 1 = yes)	0.4106	0.2072	.0483
Transportation (0 = paramedics, 1 = other)	0.9926	0.2278	.0001
$R^2 = .0581$			

[a]Time between symptom onset and decision to seek medical help.
[b]Time between decision to seek medical help and arrival at medical facility.

This study has limited generalizability because it reports the behavior of persons who successfully obtained emergency medical care. Therefore, this study does not enhance our knowledge about persons who decided not to seek care, who sought care elsewhere, or who did not survive long enough to be hospitalized. Additional limitations include sample bias resulting from the exclusion of patients seen in the emergency room but not admitted, or admitted but not interviewed before discharge; measurement error with respect to self-reported data on duration of care-seeking behavior and patients' perceptions, motivations, and behaviors; and the possibility of a chance finding of a significant association owing to the large number of comparisons made.

The results of this study emphasize the value of measuring multiple dimensions of SES,[37] including structural access to medical care, and underscore the value of conducting within-group studies of African Americans.[19] The care-seeking behavior of African Americans in our sample appears to have been influenced by factors similar to those found to influence the behavior of patients from other cultural groups, including symptom progression, degree of incapacitation, and patients' involvement of others in the decision to obtain emergency care and in arranging travel to the hospital. Assurance of access to a source of medical care as a result of HMO membership also appears to influence the time African Americans spend in deciding to seek emergency care for acute chest pain. □

References

1. Gillum RF. Coronary heart disease in Black populations, I: mortality and morbidity. *Am Heart J.* 1982;104:839–851.
2. *Report of the Secretary's Task Force on Black and Minority Health. Vol. 4: Cardiovascular and Cerebrovascular Disease.* Washington, DC: US Dept of Health and Human Services; 1986;620–638,40716.
3. Chapman JM, Frerichs RR. *Cardiovascular Disease in Los Angeles 1970–79.* Los Angeles, Calif: American Heart Association/Greater Los Angeles Affiliate; July 1982.
4. Haywood LJ. Coronary heart disease mortality/morbidity and risk in Blacks, I: clinical manifestations and diagnostic criteria: the experience with the Beta Blocker Heart Attack Trial. *Am Heart J.* 1984;108:787–793.
5. Toffler GH, Stone PH, Muller JE, et al. Effects of gender and race on prognosis after myocardial infarction: adverse prognosis for women, particularly Black women. *J Am Coll Cardiol.* 1987;9:473–482.
6. Sempos C, Cooper R, Kovar MG, McMillen M. Divergence of the recent trends in coronary mortality for the four major race-sex groups in the United States. *Am J Public Health.* 1988;78:1422–1427.
7. Wing S, Casper M, Riggan W, Hayes C, Tyroler HA. Socioenvironmental characteristics associated with the onset of decline of ischemic heart disease mortality in the United States. *Am J Public Health.* 1988;78:923–926.
8. Polednak AP. Mortality from diabetes mellitus, ischemic heart disease, and cerebrovascular disease among blacks in a higher income area. *Public Health Rep.* 1990;105:393–399.
9. Keil JE, Loadholt CB, Weinrich MC, Sandifer SH, Boyle E Jr. Incidence of coronary heart disease in Blacks in Charleston, South Carolina. *Am Heart J.* 1984;108:779–786.
10. Keil JE, Saunders DE, Lackland DT, et al. Acute myocardial infarction: period prevalence, case fatality, and comparison of black and white cases in urban and rural areas of South Carolina. *Am Heart J.* 1985;109:776–784.
11. Curry CL. Coronary artery disease in African-Americans. *Circulation.* 1991;83:1474–1475.
12. Strogatz DS. Use of medical care for chest pain: differences between Blacks and Whites. *Am J Public Health.* 1990;80:290–294.
13. Schmidt SB, Borsch MA. The prehospital phase of acute myocardial infarction in the era of thrombolysis. *Am J Cardiol.* 1990;65:1411–1415.
14. Oberman A, Cutter G. Issues in the natural history and treatment of coronary heart disease in Black populations: surgical treatment. *Am Heart J.* 1984;108:688–694.
15. Maynard C, Fisher LD, Passamani ER, et al. Blacks in the Coronary Artery Surgery Study (CASS): race and clinical decision making. *Am J Public Health.* 1986;76:1446–1448.
16. Wenneker MB, Epstein AM. Racial inequalities in the use of procedures for patients with ischemic heart disease in Massachusetts. *JAMA.* 1989;261:253–257.
17. Ford E, Cooper R, Castaner A, et al. Coronary arteriography and coronary bypass surgery among Whites and other racial groups relative to hospital-based incidence rates for coronary artery disease: findings from NHDS. *Am J Public Health.* 1989;79:437–440.
18. Cooper RS, Simmons B, Castaner A, et al. Survival rates and prehospital delay during myocardial infarction among Black persons. *Am J Cardiol.* 1986;57:208–211.
19. Jackson JS. Growing old in Black America: research on aging black populations. In: Jackson JS, ed. *The Black American Elderly: Research on Physical and Psychosocial Health.* New York, NY: Springer Publishing Co; 1988:3–16.
20. Haan M, Kaplan G. Poverty and health: prospective evidence from the Alameda County Study. *Am J Epidemiol.* 1987;125:989–998.
21. James SA. Coronary heart disease in black Americans: suggestions for research on psychosocial factors. *Am Heart J.* 1984;108:833–838.
22. Keil JE, Sutherland MS, Knapp RG, Tyroler HA. Does equal socioeconomic status in black and white men mean equal risk of mortality? *Am J Public Health.* 1992;82:1133–1136.
23. Berkanovic E, Telesky C. Mexican-American, Black-American and White-American differences in reporting illnesses, disability and physician visits for illnesses. *Soc Sci Med.* 1985;20:567–577.
24. James SA, Wagner EH, Strogatz DS, et al. The Edgecombe County high blood pressure control program, II: barriers to the use of medical care among hypertensives. *Am J Public Health.* 1984;74:468–472.
25. Alonzo AA. The mobile coronary care unit and the decision to seek medical care during acute episodes of coronary artery disease. *Med Care.* 1980;18:297–318.
26. Hollingshead AB. The two-factor index of social position. New Haven, Conn: Mimeograph; 1956.
27. Work JW, Ferguson JC, Diamond GA. Limitations of a conventional logistic regression model based on left ventricular ejection fraction in predicting coronary events after myocardial infarction. *Am J Cardiol.* 1989;64:702–707.
28. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med.* 1993;118:201–221.
29. Weissman JS, Stern R, Fielding SL, Epstein AM. Delayed access to health care: risk factors, reasons and consequences. *Ann Intern Med.* 1991;114:325–331.
30. Turi ZG, Dyonr PH, Muller JE, et al. Implications for acute intervention related to time of hospital arrival in acute myocardial infarction. *Am J Cardiol.* 1986;58:203–209.
31. Alonzo AA. The impact of the family and lay others on care-seeking during life-threatening episodes of suspected coronary artery disease. *Soc Sci Med.* 1986;22:1297–1311.
32. Hackett TP, Cassem NH. Factors contributing to delay in responding to the signs and symptoms of acute myocardial infarction. *Am J Cardiol.* 1969;24:651–658.
33. Mitic WR, Perking J. The effect of a media campaign on heart attack delay and decision times. *Can J Public Health.* 1984;75:414–418.
34. Frank E, Winkleby M, Fortmann SP, Farquhar JW. Cardiovascular disease risk factors: improvements in knowledge and behavior in the 1980s. *Am J Public Health.* 1993;83:590–593.
35. Herlitz J, Hartford M, Blohm M, et al. Effect of a media campaign on delay times and ambulance use in suspected acute myocardial infarction. *Am J Cardiol.* 1989;64:90–93.
36. Alonzo AA. The impact of physician consultation on care-seeking during acute episodes of coronary heart disease. *Med Care.* 1977;15:34–50.
37. Winkleby MA, Jatulis DE, Frank E, Fortmann SP. Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health.* 1992;82:816–820.

Behavioral Risk Factors: A Comparison of Latinos and Non-Latino Whites in San Francisco

ABSTRACT

Objectives. The purpose of the study was to evaluate differences between Latino and non-Latino White adults in health-related behavioral risk factors.

Methods. Telephone interviews were conducted with 652 Latinos and 584 non-Latino Whites in San Francisco selected by random-digit dialing.

Results. Latino men and women, compared with their non-Latino White counterparts, were less likely to have consumed any alcoholic beverage in the previous month (59% and 29% vs 77% and 75%, respectively), consumed fewer drinks per week (6.6 and 3.0 vs 8.9 and 5.1, respectively), and were more likely to be sedentary (40% and 46% vs 17% and 23%). Latina women were less likely than non-Latina Whites to smoke cigarettes (8% vs 29%), to have ever had a Pap smear (76% vs 93%), and to have ever had a clinical breast examination (81% vs 96%). Multivariate analyses adjusting for sex, age, education, and employment confirmed univariate findings.

Conclusions. Behavioral risk factor profiles by ethnicity help emphasize priorities of health promotion programs for a community. Latino needs include maintenance of limited consumption of alcohol and cigarettes, promotion of regular physical activity, and increasing use of low-cost cervical and breast cancer screening tests. (*Am J Public Health.* 1994;84:971–976)

Eliseo J. Pérez-Stable, MD, Gerardo Marín, PhD, and Barbara VanOss Marín, PhD

Introduction

Since 1981 the Centers for Disease Control and Prevention (CDC) and collaborating state health departments have been conducting telephone surveys about health behaviors. The surveys have evolved into the Behavioral Risk Factor Surveillance System, which now provides state-specific estimates of health risk behaviors and delivery of clinical preventive services for 45 states and the District of Columbia.[1] These data are used to monitor trends in behavior changes and in planning programs or legislation at the local or state level. Comparisons of health risk behaviors between Whites and African Americans have been reported,[2] and a survey in New York State compared Whites, African Americans, and Latinos on behavioral risk factors for cardiovascular disease.[3] Other studies including large samples of Latinos have not been published.

Regional and national data on selected behavioral risk factors among Latinos, such as cigarette smoking, have been studied.[4–7] In a recent national survey, the overall prevalence of cigarette smoking among US Latinos was 23%, compared with 25.6% for Whites and 26.2% for African Americans; however, this rate reflects a substantially lower smoking rate among Latina women (16.3%) than among Latino men (30.9%).[8] With regard to alcohol use, Latinos have been found to have a pattern of higher abstention rates overall, but Latino men have been found to be more likely than others to use alcohol in excess, especially in binge patterns.[9] In addition, national mortality data show that Latinos are more likely than non-Latino Whites to die from chronic liver disease.[10] Results from the Hispanic Health and Nutrition Examination Survey (HHANES) have shown that, compared with non-Latino Whites, Latina women have a two- to threefold increase in the prevalence of overweight[11] and that Latinos overall have a two- to fourfold increase in the prevalence of diabetes[12] and a similar rate of hypertension.[13] Data from two regional studies indicate that Mexican Americans lead more sedentary lives than non-Latino Whites.[14,15] Finally, the HHANES and other surveys have found that 35% of Mexican Americans, 28% of Cuban Americans, and 22% of Puerto Ricans in the United States do not have any form of health insurance, compared with 10% of Whites and 18% of African Americans.[16] Thus it is not surprising that Latinos report having had fewer medical examinations or recommended cancer screening tests.[17]

To determine local priorities in public health education for Latinos and other ethnic/racial groups, systematic needs assessments are required. We conducted a community survey among Latinos and non-Latino Whites living in the same census tracts of San Francisco to identify differences that may exist in the prevalence of selected behavioral risk factors.

Eliseo J. Pérez-Stable is with the Division of General Internal Medicine, Department of Medicine, and the Medical Effectiveness Research Center for Diverse Populations, University of California, San Francisco. Gerardo Marín is with the Department of Psychology, University of San Francisco. Barbara VanOss Marín is with the Department of Epidemiology and Biostatistics, University of California, San Francisco.

Requests for reprints should be sent to Eliseo J. Pérez-Stable, MD, 400 Parnassus Ave, Room A-405, Box 0320, San Francisco, CA 94143-0320.

This paper was accepted September 23, 1993.

TABLE 1—Demographic Characteristics of Latino and Non-Latino Whites Interviewed for Behavioral Risk Factor Profile, San Francisco, 1989

	Latinos (n = 652)		Non-Latino Whites (n = 584)	
	No.	%	No.	%
Sex				
Male	274	42	291	50
Female	378	58	293	50
Age				
18–29 y	211	32	137	23
30–34 y	127	20	98	17
35–44 y	143	22	179	31
45–65 y	171	26	170	29
Mean age, y (SD)	37 (13)		39 (12)	
Education				
0–6 y	132	20	2	<1
7–11 y	150	23	17	3
12 y	176	27	131	22
≥13 y	183	28	429	73
Missing	11	2	5	1
Mean education, y (SD)	10.8 (4.2)		15 (2.8)	
Employed	437	67	415	71

Methods

All subjects were sampled from the 50 census tracts in San Francisco County in which at least 10% of the population was reported to be Latino in the 1980 US census.[18] This sampling universe represented 67% (n = 55 541) of all Latinos (n = 83 373) and 29% (n = 98 250) of all non-Latino Whites (n = 337 118) living in San Francisco in 1980.[18] Telephone prefixes corresponding to the census tracts were identified by means of a reverse telephone directory. A modified Mitofsky-Waksberg[19] method for random-digit dialing was applied to identify eligible respondents.[20]

A household was considered to be eligible if the person answering the telephone identified himself or herself as Latino/Hispanic or non-Latino Caucasian/White, or if that person identified the majority of residents in the household as belonging to one of these groups. Potential respondents were initially asked if they considered themselves or the majority of residents in the household to be Latino/Hispanic. One third of the participating Latino households were randomly chosen to respond to the 24-item behavioral risk factor questionnaire (the remainder responded to a shorter questionnaire on smoking behavior only). Within a given household, the adult between 18 and 65 years of age who had most recently celebrated a birthday was invited to respond to the survey. Potential respondents who refused to participate were not replaced within a household. There were 40 refusals (6%) among Latinos, but refusal rates were not computed for non-Latino Whites. All initial refusals were called a second time and unavailable eligible respondents were called approximately 10 times before they were counted as refusals. Respondents who classified themselves as Latino but who were subsequently found not to be Latino by their responses on family background were excluded from the study.

The questionnaire was administered anonymously by trained, supervised, and experienced bilingual and bicultural Latino interviewers of both sexes after subjects gave verbal consent. Latino participants responded in Spanish (80.8%) or in English (19.2%). The survey instrument and a more detailed description of the interview procedure are available from the authors upon request.

Sex, age, education, employment in the previous 2 weeks, birthplace, and national background (for persons not born in the United States) of respondents were obtained. Latinos also completed a five-item acculturation scale and their scores were dichotomized into less (scores of 1 to <3) and more (scores of 3 to 5) acculturated groups.[21] Smoking behavior items were similar to those of earlier surveys.[7] We selected 24 of the 55 questions (not including questions about demographics) from the 1989 Behavioral Risk Factor Survey (BRFS, provided by CDC) to include items on seat belt use, alcohol consumption, exercise, perceived weight status, and preventive practices. Part of the BRFS questionnaire was found to have acceptable to high item reliability and to be consistent among Whites, African Americans, and Latinos in New York State.[3]

On the basis of responses to the questionnaire, eight behavioral risk factors were identified as outcome (dependent) variables: (1) any use of alcohol in the previous month; (2) binge drinking (defined as having had five or more drinks on one occasion in the previous month); (3) driving a motor vehicle after having had too much to drink in the previous month; (4) current cigarette smoking; (5) infrequent use of seat belts (defined by a response of "sometimes," "seldom," or "never" to a question about how often belts were used); (6) sedentary life-style (defined as no leisure-time physical activity); (7) no routine medical checkup in the previous year; (8) never having had a blood cholesterol test. Additional behavioral risk factors for women were never having had a Pap test, never having had a clinical breast examination, and never having had a mammogram.

Descriptive data were analyzed by standard techniques. Proportions were compared by corrected chi-square tests and continuous variables were compared by analysis of variance. Proportions of behavioral risk factors were age-adjusted to the 1980 US census[18] by the direct method.[22] Ethnic differences (with non-Latino Whites as the referent) on behavioral risk factor rates were evaluated by logistic regression analyses adjusted for sex (women as referent), age (continuous), education (continuous), and employment (being employed as referent). Additional multivariate analyses of the Latino sample were adjusted for acculturation level (more acculturated as referent).

Results

The demographic characteristics of the 652 Latinos and 584 non-Latino Whites interviewed by telephone between June and August 1989 are shown in Table 1. A larger proportion of Latino respondents were women ($P < .01$) and were between 18 and 34 years of age ($P < .001$). Considerably fewer Latinos than non-

Latino Whites had graduated from high school ($P < .001$). Seventy-four percent of the Latino respondents scored as less acculturated, and 77% were born in Latin America. The foreign-born Latinos had been living in the United States for an average of 12.9 years. Thirty percent of the Latinos were of Mexican background, 53% were Central American, and 14% were from other Latin American countries or Spain.

Any Alcohol Use, Binge Drinking, and Drinking and Driving

Table 2 shows the proportions of respondents reporting the selected behavioral risk factors. A significantly lower proportion of Latinos than non-Latino Whites reported any use of alcohol in the previous month ($P < .001$ for both men and women). Latinos consumed significantly fewer drinks per week than their non-Latino White counterparts (6.6 vs 8.9 for men, 3.0 vs 5.1 for women; both Ps $< .001$). Only 6% of Latino men and 0.5% of Latina women reported averaging more than two drinks per day, compared with 16% and 6%, respectively, of their non-Latino counterparts. Although there were no differences in reported binge drinking by ethnicity, a higher proportion of Latino than White men reported drinking and driving at least once. Six percent of Latino men reported having driven a motor vehicle two or more times in the previous month when they may have had too much to drink, compared with 1.3% of non-Latino White men ($P < .01$). Of the 16 Latinos who reported this behavior, 14 were men, 14 were born in Latin America, 13 were 34 years of age or younger, 11 had less than a high school education, and 11 were employed.

Smoking Behavior

The prevalence of smoking among Latina women was very low, although rates for men were similar for both ethnic groups. Latino smokers of both sexes reported smoking fewer than half as many cigarettes per day as non-Latino Whites ($P < .001$ for men, $P = .01$ for women).

Seat Belt Use, Perceived Overweight Status, and Sedentary Life-Style

The proportions of Latinos and non-Latino Whites who reported infrequent use of seat belts were similar (Table 2). However, 5 of the 36 Latinos, compared with none of the 40 non-Latino Whites, who *never* used seat belts reported driving at least twice when they may have had too much to drink.

The proportions of respondents who perceived themselves as overweight were similar for both ethnic groups (for men, 29% of Latinos and 25% of non-Latino Whites; for women, 46% of Latinas and 38% of non-Latina Whites). Latinos of both sexes were significantly less likely than non-Latino Whites to ever exercise and to report exercising daily. In both ethnic groups, similar proportions of self-perceived overweight respondents reported an exercise frequency of at least three times per week.

Medical Examinations and Cancer Screening Tests

After results were stratified by sex, similar proportions of Latinos and non-Latino Whites reported having visited a physician for a routine checkup in the previous year, having ever had a cholesterol blood test (Table 2), and having been told by a physician that they had diabetes or hypertension. Among respondents who were 45 years of age or older, fecal occult blood tests in the previous year were reported less often by Latinos than by their non-Latino counterparts (for men, 34% vs 53%; for women, 49% vs 62%). Latina women reported fewer lifetime Pap smears and clinical breast examinations than non-Latina Whites, and similar proportions of women in each ethnic group reported ever having had a mammogram (Table 2).

TABLE 2—Percentages[a] of Behavioral Risk Factors among Latino and Non-Latino Whites Interviewed, San Francisco, 1989

	Men		Women			
	Latino (n = 274)	White (n = 291)	Latina (n = 378)	White (n = 293)	OR[b]	95% CI
Used alcohol in previous month	59	77	29	75	0.31	0.23, 0.42
Binge drinking	33	32	11	17	0.81	0.58, 1.13
Drinking and driving	9	6	2	4	1.50	0.97, 2.20
Current smoking	26	30	8	29	0.37	0.26, 0.52
Mean cigarettes/day	6.8	18.5	6.6	16.1		
Infrequent use of seat belts	25	25	24	18	0.83	0.60, 1.16
Sedentary life-style	40	17	47	23	2.15	1.59, 2.91
No physician visit in previous year	34	37	16	18	0.71	0.52, 0.98
No cholesterol test ever	13	15	10	15	0.95	0.71, 1.26
No Pap smear ever	...	...	24	7	2.41	1.30, 4.45
No clinical breast exam ever	...	...	19	4	3.09	1.45, 6.59
No mammogram ever[c]	...	...	26	23	0.88	0.58, 1.34

[a]All proportions are age-adjusted to the 1980 US census by the direct method.
[b]Odds ratio of Latinos reporting the behavioral risk factor compared with non-Latino Whites, adjusted for sex where appropriate, education, age, and employment.
[c]Only for women aged 35 years and older (n = 146 Latinas and 138 Whites).

Multivariate Analyses: Correlates of Behavioral Risk Factors

Multivariate logistic regression analyses showed that Latino ethnicity was a significant predictor for selected behaviors after adjustments were made for sex, education, age, and employment (Table 2). Latinos were more likely to abstain from alcohol (odds ratio [OR] = 3.6, 95% confidence interval [CI] = 2.7, 4.8; $P < .001$) and less likely to have consumed any alcoholic beverage in the previous month (OR = 0.31, 95% CI = 0.23, 0.42; $P < .001$), but none of the other alcohol-related behaviors were significantly associated with ethnicity. Latinos were also less likely to smoke cigarettes (OR = 0.37, 95% CI = 0.26, 0.52; $P < .001$) and more likely to be sedentary (OR = 2.15, 95% CI = 1.59, 2.91; $P < .001$). Finally, Latinas were less likely than non-Latina Whites to report ever having had a Pap smear (OR = 2.41, 95% CI = 1.30, 4.45; $P = .005$) or a clinical breast examination (OR = 3.09, 95% CI = 1.45, 6.59; $P = .003$).

Other demographic variables predicted behaviors independently of ethnicity. Men were more likely to have consumed any alcohol in the previous month (OR = 2.1, 95% CI = 1.7, 2.6), more likely to consume alcohol in a binge

TABLE 3—Percentages[a] of Behavioral Risk Factors among Latino Men and Women Interviewed, by Acculturation Score, San Francisco, 1989

	Men		Women			
	Score <3 (n = 202)	Score 3–5 (n = 72)	Score <3 (n = 280)	Score 3–5 (n = 97)	OR[b]	95% CI
Used alcohol in previous month	54	72	22	46	0.41	0.27, 0.62
Binge drinking	35	28	9	17	0.83	0.51, 1.35
Drinking and driving	10	8	2	2	1.04	0.39, 2.79
Current smoking	26	27	7	12	0.61	0.36, 1.03
Mean cigarettes/day	8	10	6	8		
Infrequent use of seat belts	28	15	22	21	1.26	0.79, 2.03
Sedentary life-style	44	30	55	29	2.05	1.35, 3.12
No physician visit in previous year	38	25	13	19	1.25	0.78, 2.01
No cholesterol test ever	13	15	8	12	1.54	1.01, 2.34
No Pap smear ever	...	...	24	7	0.69	0.36, 1.33
No clinical breast exam ever	...	...	19	4	1.88	0.85, 4.15
No mammogram ever[c]	...	...	26	23	0.84	0.48, 1.46

[a]All proportions are age-adjusted to the 1980 US census by the direct method.
[b]Odds ratio of less acculturated (score <3) Latinos reporting the behavioral risk factor compared with more acculturated (score 3–5) Latinos, adjusted for sex where appropriate, education, age, and employment.
[c]Only for women aged 35 years and older (n = 146).

pattern (OR = 3.3, 95% CI = 2.5, 4.5), more likely to drink and drive (OR = 2.7, 95% CI = 1.0, 4.0), less likely to use seat belts frequently (OR = 1.34, 95% CI = 1.02, 1.80), and less likely to be sedentary (OR = 0.75, 95% CI = 0.57, 0.97). Persons with less than a high school education were more likely to abstain from alcohol (OR = 1.4, 95% CI = 1.1, 1.9), more likely to drink and drive (OR = 1.96, 95% CI = 1.3, 2.9), more likely to smoke cigarettes (OR = 1.5, 95% CI = 1.1, 1.9), less likely to use seat belts (OR = 2.2, 95% CI = 1.6, 3.1), and more likely to be sedentary (OR = 2.6, 95% CI = 1.96, 3.50). Persons 35 to 65 years of age were less likely to smoke cigarettes (OR = 0.60, 95% CI = 0.46, 0.76) and more likely to be sedentary (OR = 1.5, 95% CI = 1.1, 1.9); women in this age group were more likely to have had at least one Pap test (OR = 2.6, 95% CI = 1.6, 4.2), more likely to have had at least one breast examination (OR = 1.8, 95% CI = 1.1, 3.0), and more likely to have had at least one mammogram (OR = 7.1, 95% CI = 5.0, 10.0).

Acculturation and Behavioral Risk Factors among Latinos

The effects of acculturation on selected self-reported behaviors among Latinos are shown in Table 3. A lower proportion of less acculturated men ($P = .01$) and women ($P < .001$) reported consuming any alcoholic beverage in the previous month. There were no significant differences by acculturation for either sex in proportions of heavy drinking, binge drinking, drinking and driving, current smoking, or seat belt use. Less acculturated men ($P = .003$) and women ($P < .001$) were more likely to be sedentary. Less acculturated men, but not women, were less likely to have had a routine checkup in the previous year ($P = .009$).

Multivariate analyses showed that less acculturated Latinos were less likely to have consumed alcoholic beverages in the previous month (OR = 0.41, 95% CI = 0.27, 0.62) and to be sedentary (OR = 2.05, 95% CI = 1.35, 3.12). Male sex was the only significant predictor of cigarette smoking among Latinos (OR = 4.2, 95% CI = 3.8, 4.6), although less acculturation was nearly significant (p = .06).

Discussion

To effectively promote healthy behaviors among ethnic and racial groups in the United States, programs need to be based on the specific cultural and behavioral characteristics of each group.[23] This study contrasts the behavioral risk factor profiles of Latinos and non-Latino Whites in San Francisco and provides information on which to base local programs. Compared with a statewide sample in California, this San Francisco sample reported somewhat higher prevalences of binge drinking, drinking and driving, and seat belt nonuse.[24] The proportions that reported smoking cigarettes, having had their cholesterol measured, and having had a mammogram were similar to those in the statewide sample.[24] Only a few substantial differences in behavioral risk factors between ethnic groups were found in our study: these differences were in the consumption of alcohol in the previous month, current smoking, sedentary lifestyle, and having had selected cancer screening tests.

The proportion of abstainers from alcohol was higher and the mean number of alcoholic drinks consumed per week was lower for both Latino men and Latina women. In an analysis of HHANES data, Markides and colleagues found that marriage was the principal demographic factor that consistently predicted less alcohol consumption among Mexican-American men.[25] Although we did not find higher rates of alcohol consumption among Latino men than among non-Latino Whites, our data indicate that a subset of less acculturated Latino men are more likely than non-Latino Whites to combine excessive drinking and driving. However, driving without a seat belt after having had too much to drink was an infrequent high-risk behavior in both ethnic groups, and the proportion reporting this behavior was similar to the 6.3% found among all men in California.[24] In the HHANES, Latinos with more years of education, higher incomes, and higher levels of acculturation reported higher alcohol consumption.[25]

In contrast to the men, 70% of Latina women reported no alcohol use in the previous month, and only 3% averaged one or more drinks per day. Marks and colleagues also found low alcohol use among women from the three major Latino subgroups participating in the HHANES, especially among the less acculturated. This protective effect of traditional Latino culture may be an important element to incorporate into comprehensive health promotion campaigns focused on alcohol use. Markides et al. reported that increasing acculturation was associated with increased alcohol intake among Mexican-American women 20 to 39 years of age.[25] We found that less acculturated Latinos of both sexes were

less likely to consume alcoholic beverages, but the proportion of heavy drinkers was not related to acculturation.

The prevalence of sedentary life-style, defined as no leisure-time physical activity in the previous week, was higher among Latino men and women than among non-Latino Whites. The possible health advantages provided by regular physical activity include a lower likelihood of diabetes.[27] Given that diabetes is two to four times more prevalent among Mexican Americans than among Whites or African Americans,[12] the failure to realize the potential benefits of regular exercise may be of particular importance for Latinos.

Latinos were less likely than non-Latino Whites to report having received a medical examination in the previous year. Differences between ethnic groups were observed for fecal occult blood tests, Pap tests, and clinical breast examinations, but not for mammograms. We did not ask about health insurance status in our survey, and it is likely that these differences reflect the lower coverage rates that have been reported for Latinos.[16] Analyses of the HHANES data have shown that access to health care was the most important predictor of use of preventive health measures for the three Latino subgroups surveyed.[28] Although there is controversy over the roles of regular medical examinations and annual fecal occult blood tests, there is no question about the benefits of screening for cervical[29] and breast[30] cancers. The proportions of Latina women in our sample who had ever had a Pap smear and who had ever had a clinical breast examination were similar to the proportions reported in the HHANES and other surveys and were considerably lower than those found for non-Latina White women.[17,28]

Although the main objective of this study was to identify ethnic differences in behavioral risk factors, the importance of years of formal education was also highlighted by our data. We found more risk behaviors associated with fewer years of educational attainment, including drinking and driving, cigarette smoking, sedentary life-style, and nonuse of seat belts. Thus, we observed differences in risk behaviors by educational attainment and by ethnicity that were independent of each other. The different patterns of risk profiles by years of education within ethnic groups may have implications in the development of health promotion campaigns.[31]

There are several limitations of this study that should be considered in interpreting the results. First, responses to an interview administered over the telephone may not reflect actual behavior, and no validation of self-reports was conducted. A recent study found that comparable estimates were produced by the BRFS and the Stanford Five-City Project Survey (supplemented by physiological data) for selected cardiovascular risk measures, but significant differences were found for measures of body weight.[32] Second, subjects were eligible for this study if they identified themselves as belonging to one of the two ethnic groups and if they lived in a household with a telephone (92% to 96% of San Francisco households have telephones). Variations in self-categorization of ethnicity and in sample selection by random-digit dialing may have affected the results. Third, a comparison of Latino respondents with non-Latino Whites who live in the same census tracts may minimize potential differences in behavioral risk factors influenced by socioeconomic status. Finally, the variables identified as predicting behavioral risk factors need to be interpreted with caution because these results are based on a single cross-sectional survey.

The differences in behavioral risk factor profiles between Latinos and non-Latino Whites emphasize the importance of ethnic group–specific priorities. This is of particular importance in California, where more than 35% of the population is non-White.[33] These data indicate that health promotion programs for San Francisco Latinos should encourage Latinos to maintain low consumption of alcoholic beverages and cigarettes, should educate less acculturated Latino men about the risks of driving a motor vehicle while intoxicated, should promote more physical activity among Latinos of both sexes, and should facilitate the use of low-cost cervical and breast cancer screening tests for Latina women. It is important to conduct health promotion campaigns in Spanish, to use easily understood messages that are culturally appropriate, and to base these messages on research that identifies attitudes and values specific to the target audience. To further define health promotion needs, minority groups in selected states should be oversampled in behavioral risk factor surveys. □

Acknowledgments

This research was supported in part by Public Health Service grant CA39260 awarded by the National Cancer Institute and by grant HS07373-01 from the Agency for Health Care Policy and Research. Dr Pérez-Stable is a Henry J. Kaiser Family Foundation Faculty Scholar in general internal medicine.

The authors acknowledge Samuel Posner and Raymond J. Gamba for assistance in data management and Rosa Marcano for supervising the interviewers for the survey.

References

1. Remington PL, Smith MY, Williamson DF, Anda RF, Gentry EM, Hogelin GC. Design, characteristics, and usefulness of state-based behavioral risk factor surveillance: 1981–87. *Public Health Rep.* 1988;103:366–375.
2. Remington PL, Forman MR, Gentry EM, Marks JS, Hogelin GC, Trowbridge FL. Current smoking trends in the United States: the 1981–1983 Behavioral Risk Factors Surveys. *JAMA.* 1985;253:2975–2978.
3. Shea S, Stein AD, Lantigua R, Basch CE. Reliability of the Behavioral Risk Factor Survey in a triethnic population. *Am J Epidemiol.* 1991;133:489–500.
4. Escobedo LG, Remington PL. Birth cohort analysis of prevalence of cigarette smoking among Hispanics in the United States. *JAMA.* 1989;261:66–71.
5. Haynes SG, Harvey C, Montes H, Nickens H, Cohen BH. Patterns of cigarette smoking among Hispanics in the United States: results from HHANES 1982–84. *Am J Public Health.* 1990;80(suppl):47–53.
6. Marcus AC, Crane LA. Smoking behavior among US Latinos: an emerging challenge for public health. *Am J Public Health.* 1985;75:169–172.
7. Marín G, Pérez-Stable EJ, Marín BV. Cigarette smoking among San Francisco Hispanics: the role of acculturation and gender. *Am J Public Health.* 1989;79:196–198.
8. Centers for Disease Control. Cigarette smoking among adults—United States, 1990. *MMWR.* 1992;41:354–362.
9. Caetano R. Drinking patterns and alcohol problems among Hispanics in the U.S.: a review. *Drug Alcohol Depend.* 1983;18:1–15.
10. Maurer JD, Rosenberg HM, Keemer JB. Deaths of Hispanic origin, 15 reporting states, 1979–1981. *Vital Health Stat [20].* 1990; no. 18:1–73. DHHS publication PHS 91–1855.
11. Centers for Disease Control. Prevalence of overweight for Hispanics—United States, 1982–1984. *MMWR.* 1989;38:838–842.
12. Flegal KM, Ezzati TM, Harris MI, et al. Prevalence of diabetes in Mexican Americans, Cubans, and Puerto Ricans from the Hispanic Health and Nutrition Examination Survey, 1982–1984. *Diabetes Care.* 1991;14(suppl 3):628–638.
13. Pappas G, Gergen PJ, Carroll M. Hypertension prevalence and the status of awareness, treatment, and control in the Hispanic Health and Nutrition Examination Survey (HHANES), 1982–84. *Am J Public Health.* 1990;80:1431–1436.
14. Hazuda JP, Stern MP, Gaskill SP, Haffner SM, Gardner LI. Ethnic differences in

health knowledge and behaviors related to the prevention and treatment of coronary heart disease: the San Antonio heart study. *Am J Epidemiol.* 1983;117:717–728.
15. Burchfiel CM, Hamman RF, Marshall JA, Baxter J, Kahn LB, Amirani JJ. Cardiovascular risk factors and impaired glucose tolerance: the San Luis Valley Diabetes Study. *Am J Epidemiol.* 1990;131:57–70.
16. Treviño FM, Moyer ME, Valdez B, Stroup-Benham CA. Health insurance coverage and utilization of health services by Mexican Americans, Mainland Puerto Ricans, and Cuban Americans. *JAMA.* 1991;265: 233–237.
17. Harlan LC, Bernstein AB, Kessler LG. Cervical cancer screening: who is not screened and why? *Am J Public Health.* 1991;81:885–890.
18. *1980 Census of Population: Age, Sex, Race, and Spanish Origin of the Population by Regions, Division and States.* Washington, DC: US Dept of Commerce, Bureau of the Census; 1981. PC80-S1-1 ed.
19. Waksberg J. Sampling methods for random digit dialing. *J Am Stat Assoc.* 1978;73:40–46.
20. Marín G, Marín BV, Pérez-Stable EJ. Feasibility of a telephone survey to study a minority community: Hispanics in San Francisco. *Am J Public Health.* 1990;80:323–326.
21. Marín G, Sabogal F, Marín BV, Otero-Sabogal R, Pérez-Stable EJ. Development of a short acculturation scale for Hispanics. *Hispanic J Behav Sci.* 1987;9:183–205.
22. Fleiss J. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley & Sons Inc; 1981.
23. Marín G, Marín BV. Research with Hispanic populations. In: Bickman L, Rog DJ, ed. *Applied Social Research Methods.* Vol. 23. Newbury Park, Calif: Sage Publications; 1991:130.
24. Anda RF, Waller MN, Wooten KG, Mast EE, Escobedo LG, Sanderson LM. Behavioral risk factor surveillance, 1988. *MMWR.* 1990;39(SS-2):1–21.
25. Markides KS, Ray LA, Stroup-Benham CA, Treviño F. Acculturation and alcohol consumption in the Mexican American population of the Southwestern United States: findings from HHANES 1982–84. *Am J Public Health.* 1990;80(suppl):42–46.
26. Marks G, Garcia M, Solis JM. Health risk behaviors of Hispanics in the United States: findings from HHANES, 1982–84. *Am J Public Health.* 1990;80(suppl):20–26.
27. Helmrich SP, Ragland DR, Leung RW, Paffenbarger RS. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *N Engl J Med.* 1991;325:147–152.
28. Solis JM, Marks G, Garcia M, Shelton D. Acculturation, access to care, and use of preventive services by Hispanics: findings from HHANES 1982–84. *Am J Public Health.* 1990;80(suppl):11–19.
29. Guzick DS. Efficacy of screening for cervical cancer: a review. *Am J Public Health.* 1978;68:125–134.
30. Shapiro S. Evidence on screening for breast cancer from a randomized trial. *Cancer.* 1977;39:2772–2782.
31. Shea S, Stein AD, Basch CE, et al. Independent associations of educational attainment and ethnicity with behavioral risk factors for cardiovascular disease. *Am J Epidemiol.* 1991;134:567–582.
32. Jackson C, Jatulis DE, Fortmann SP. The Behavioral Risk Factor Survey and the Stanford Five-City Project Survey: A comparison of cardiovascular risk estimates. *Am J Public Health.* 1992;82:412–416.
33. *1990 Census of Population and Housing. Summary Population and Housing Characteristics, California.* Washington, DC: US Dept of Commerce, Economics and Statistics Administration; 1991.

US Public Health Service/HCFA Conference on Primary Care to Be Held in September

The US Public Health Service and the Health Care Financing Administration of the Department of Health and Human Services will convene the Second National Primary Care Conference on September 11 through 13, 1994, at the Fairmont Hotel in Dallas, Tex. The theme of the conference is "Making Primary Care Work Under Health Care Reform." The 2-day conference of panel discussions and plenary and break-out sessions will be attended by primary care and other health care practitioners; national, state, and local policymakers; health care administrators; researchers; educators; and consumers.

Unlike the First National Primary Care Conference that was convened 2 years ago in Washington, DC, and that centered on policy related issues, this conference will address very practical issues and questions. It will provide a national forum that focuses on the primary care changes that are already occurring under state health care reform as they relate to primary care systems, workforce issues, and needs of the underserved. Actual case studies will be discussed with decisionmakers from health care provider organizations, academic health centers, and both the legislative and executive branches of government.

Cosponsors of the conference include the W. K. Kellogg Foundation, the Robert Wood Johnson Foundation, the Josiah Macy Jr. Foundation, the National Academy for State Health Policy, the Henry J. Kaiser Family Foundation, and the Pew Charitable Trusts.

A $250 registration fee covers all conference costs, including materials, meals, reception, and special events. For additional information and to ensure that you receive the program brochure, contact Rose Salton, Social & Scientific Systems, 7101 Wisconsin Ave, Suite 1300, Bethesda, MD 20814; tel (301) 986-4870; fax (301) 913-0351.

Social Influences on the Sexual Behavior of Youth at Risk for HIV Exposure

ABSTRACT

Objectives. Adolescents are increasingly at risk for infection with human immunodeficiency virus (HIV) and other sexually transmitted diseases, especially in poor urban minority communities. To aid the design of interventions in these communities, this study investigated the role of knowledge, attitudes, perceived parental monitoring, and peer behavior in the onset and progression of sexual behavior in children at risk for exposure to HIV.

Methods. A computerized personal interview was administered to 300 African-American 9- to 15-year-old children living in six public housing developments in a large US city.

Results. Although children's knowledge about the hazards of sex increased with age, their sexual activity also increased (from 12% sexually experienced at 9 years of age to more than 80% experienced at 15 years of age). Parental monitoring appeared able to influence sexual activity. However, the perceived behavior of friends was associated with the rate at which sexual activity progressed with age and the degree to which condom use was maintained with age.

Conclusions. The early onset and prevalence of sexual behavior and the importance of peer group influence call for early interventions that simultaneously influence the parents and peers in children's social networks. (*Am J Public Health.* 1994; 84:977–985)

Daniel Romer, PhD, Maureen Black, PhD, Izabel Ricardo, PhD, Susan Feigelman, MD, Linda Kaljee, PhD, Jennifer Galbraith, MA, Rodney Nesbit, BA, Robert C. Hornik, PhD, and Bonita Stanton, MD

Introduction

Human immunodeficiency virus (HIV) infection is increasingly spread by heterosexual transmission in urban minority communities of the United States and is now the sixth leading cause of death in young people.[1] Large surveys of minority youth suggest that sexual activity is initiated at an early age,[2,3] that protective measures such as condoms are not consistently used, and that use of protective measures may decline with age.[4,5] Minority youths living in poverty are at particularly great risk in view of their high rates of sexually transmitted diseases and adolescent pregnancies.[6]

In the absence of effective medical interventions, educational programs designed to delay the onset of sexual contact and to encourage the use of condoms when contact occurs are needed to protect young people from exposure to HIV as well as other sexually transmitted diseases.[1,7] The present research was undertaken to aid in the design of interventions with high-risk children during the ages when sexual practices are first developing (years 9 through 15).

Most educational interventions to prevent HIV infection focus on the health-related attitudes and behavioral skills of the *individual* as program goals. Although these factors are important in initiating safe behavior, they may not be sufficient to maintain this behavior when social environments continue to encourage risk.[8,9] Social environments are especially critical during the early adolescent years when standards for behavior are often set by parents and peers.[10] Even if individual attitudes and skills support safe behavior, social norms can influence the onset and progression of risky behavior.[5,11,12] Although research suggests that sexual expression, especially in boys, is greatly influenced by hormonal changes that occur naturally during maturation,[13,14] perceived peer behavior may still define boundaries for the expression of hormonal influence.[15]

Parents and adults also play a major role in socialization and control a good deal of children's activities. Parenting styles are related to several aspects of childhood development,[16,17] including risky sexual behavior[18] and delinquency.[19] Parental supervision and monitoring may be particularly relevant to sexual expression by communicating parents' concern about risky behavior and raising children's awareness of the costs of such behavior.

It is critical to understand the social processes that influence sexual expression so that prevention programs can be designed to incorporate these influences.[20] The present research was designed to provide an understanding of the social influences of parents and peers on the initiation and adoption of sexual practices in a population at great risk for exposure to sexually transmitted diseases and HIV.

Daniel Romer and Robert C. Hornik are with the Center for International, Health, and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. Maureen Black, Izabel Ricardo, Susan Feigelman, Linda Kaljee, Jennifer Galbraith, Rodney Nesbit, and Bonita Stanton are with the Department of Pediatrics, University of Maryland Medical School, Baltimore.

Requests for reprints should be sent to Daniel Romer, PhD, Annenberg School for Communication, University of Pennsylvania, 3620 Walnut St, Philadelphia, PA 19104.

This paper was accepted September 1, 1993.

Methods

A survey of children 9 to 15 years of age was conducted from March to May 1992 in six public housing sites in a large US city. The survey was conducted in a residential setting (rather than in schools or clinics) to increase the chances of reaching high-risk children. Public housing was selected because it ensured that residents would be close to or below official poverty levels. The present sites contained approximately 4000 households with approximately 1600 children between the ages of 9 and 15 years. Virtually all of the households were African American, more than 80% were headed by women, and all received some form of publicly financed assistance.

Prior to conducting the survey, project staff met with public housing officials and resident councils to secure their endorsement of the survey and cooperation in recruiting respondents. Children were recruited with the assistance of several adult residents and with the use of posters and letters to residents urging them to bring their children to an interview location (a public room at the site) during scheduled hours. Interviewing was conducted for about 2 weeks per site. There was a tendency to recruit more children of younger ages under the assumption that their sexual activity would be less developed and would require larger numbers to achieve stable estimates. Informed consent was obtained from both caretakers and children, and each child received $5 for participating. In order to assess the test–retest reliability of the interview, a subsample of 34 children was recruited to return about a week later to repeat the interview. The distribution of these children in age and gender was similar to that of the total sample.

The interview was conducted with a Macintosh computer that was programmed to administer the questions orally over earphones. The program provided a personalized version of the interview to take into account the child's gender, age, and level of sexual development. An adult interviewer explained the use of the computer and remained available to answer questions, but children completed the interview largely on their own. The major purpose for using a computerized interview was to permit children to respond with as much confidentiality as possible while still maintaining a high level of attention. A written questionnaire would have posed comprehension problems for younger children and those with weak reading skills. With the computerized interview, children listened to the questions privately over earphones, and only the computer could "hear" their responses (which were entered by using a mouse to press buttons visually displayed on the screen). Most of the responses were of the yes/no variety and, in all cases, were explained during the interview. The interview was self-paced; children could repeat any question that was not clear on the first presentation, and they could skip any question that they could not or did not want to answer. The entire interview took about 45 minutes.

The interview was divided into several sections, the results of three of which are reported here. The opening section focused on the child's typical activities and included an in-depth discussion of his or her friends. Children were asked to name up to nine same- and opposite-sex friends and to report on various characteristics of each friend.

The second section of the interview focused on the child's sexual behavior. Initial questions asked about kissing and then asked whether the child had ever had sex with anyone. Earlier focus groups with children across our age range indicated that they knew the meaning of this question.[21] For those children more than 10 years of age who reported having had sex, the interview went on to ask whether they had engaged in sex in the last 3 months. In addition, children were asked whether they had ever used birth control in general and condoms in particular and whether they had used a condom on the last sexual contact. Children who were 9 and 10 years of age were asked only whether they had used a condom on the last occasion. Children who reported not yet having had sex were asked whether they thought they would have sex in the next year.

On the basis of these responses, two scales of sexual practice were constructed. One was a scale of sexual activity ranging from never having had sex (0) to having had sex in the last 3 months (1.5). The scale ranged from only 0 to 1 for 9- and 10-year-old children because they were not asked about sex in the previous 3 months. We assumed that, for children of this age, any sex would be rare enough that further differentiation would be uninformative. The scale had a test–retest reliability of .86, suggesting that it was a stable indicator of sexual activity.

A second scale was constructed to assess the level of condom use (among those children who had ever engaged in sex). The scale ranged from never having used a condom (0) to having used a condom on the last sexual occasion (2). The scale had only 2 points for children 9 and 10 years of age (0 or 2); older children who reported use of condoms were further divided into those who did (2) and those who did not (1) use one on the last occasion. This scale also had a high test–retest reliability (r = .83).

To assess the influence of peers, two questions asked about the behavior of the friends who had been listed in the beginning of the interview. One question concerned the respondent's perception of how many of his or her friends "have had sex" (none, some, or most). The other question asked how many of the respondent's friends (of those whose friends had ever engaged in sex) used condoms (none, some, or most). These items also had adequate test–retest reliabilities (r's = .67 and .57, respectively).

To assess the influence of caretakers, a series of six questions addressed the frequency with which the child's parents or other responsible adults monitored the child's social behavior outside the home (never, sometimes, or always). Two questions asked whether caretakers generally knew where the child was and whether they expected a call if the child was late. The other four questions asked whether caretakers knew whom the child was with and details of the child's activities. The two sets of items were combined into separate indices of caretaker monitoring that were only moderately correlated (r = .41). The shorter, "whereabouts" scale had an internal consistency score (Cronbach's alpha) of .51, and the longer, "detailed knowledge" scale had a Cronbach's alpha of .70. These scores suggested that the scales measured consistent aspects of caretaker behavior. The more internally consistent scale was also more stable over time (r = .44 vs .33).

All children were also asked to rate, on a 5-point scale ranging from "very bad" to "very good," how they would feel if they were to have sex in the next year (test–retest r = .74). This scale was designed to measure the child's attitude toward having sex and was included as a diagnostic score to interpret the relations between social influences and sexual behavior. The final section of the interview contained a series of knowledge questions about HIV and acquired immunodeficiency syndrome (AIDS) that were adapted from tests frequently used in evaluating AIDS education.[22,23] These

brief statements were answered as true or false. Four items stated common misconceptions about the transmission and cure of HIV infection. Two items assessed the importance of using condoms to prevent transmission. Three other items referred to hazards of sex in transmitting HIV. Our interest was in assessing deficits in knowledge about HIV and in determining whether those deficits were related to sexual practices or to peer and caretaker influences. At the close of the interview, project staff gave children explanations of the correct answers to all of the questions and gave them an opportunity to ask additional questions about AIDS.

Results

Of the 316 children who were screened to participate in the survey, 300 successfully completed the interview (160 girls and 140 boys), approximately 50 respondents per housing site. Table 1 contains the numbers of children by age and gender, as well as the results for the sexual activity questions. As many as 12% of the 9-year-old children reported having had sex. By 15 years of age, all of the boys and more than 80% of the girls claimed to have had sex. Of those children who had engaged in sex, high proportions also claimed to have had sex in the last 3 months. A multiple logistic regression analysis of ever having had sex revealed that both age, $\chi^2(1) = 5.9$, $P < .05$, and gender, $\chi^2(1) = 31.8$, $P < .05$, predicted sexual activity. There was no interaction between gender and age, $\chi^2(1) = 0.08$, $P = .78$.

Figure 1 shows the sexual activity scale scores for boys and girls as a function of age. The results of a multiple regression analysis (Table 2) indicated that sexual activity increased at the same rate with age in boys as in girls but that boys had a consistently higher level of sexual activity at any age.

Table 1 also contains the percentage of children at each age and for each gender who reported ever having used a condom and having used a condom on the last occasion of sex (among sexually experienced respondents). Experience with condoms was quite high. Of the children over 10 years of age, more than 90% of the girls and more than 75% of the boys had used a condom. Rates of use for the last sexual occasion were also high. Because the number of respondents within each age group was small, we aggregated

TABLE 1—Reports of Sexual Activity and Condom Use, by Age and Gender

	Age, y																				
	Girls										Boys										
	9 (n = 32)	10 (n = 30)	11 (n = 24)	12 (n = 22)	13 (n = 25)	14 (n = 15)	15 (n = 12)	Total (n = 160)	9 (n = 20)	10 (n = 25)	11 (n = 24)	12 (n = 19)	13 (n = 21)	14 (n = 17)	15 (n = 14)	Total (n = 140)
Sexual activity																
Had sex	3.1	13.3	12.5	13.6	48.0	40.0	83.3	24.4	25.0	28.0	41.7	63.2	76.2	82.4	100.0	55.7
Last 3 months	. . .[a]	. . .[a]	8.3	0.0	28.0	26.7	58.3	20.4[b]	. . .[a]	. . .[a]	20.8	42.1	42.9	70.6	78.6	47.4[d]
Next year	6.3	19.9	20.8	31.8	8.0	20.0	8.3	17.5	25.0	24.0	16.7	26.2	14.3	5.9	0.0	17.1
Condom use (among those who had sex)																
Used condoms	. . .[a]	. . .[a]	100.0	66.7	83.3	100.0	100.0	94.1[c]	. . .[a]	. . .[a]	60.0	83.3	68.8	92.9	71.4	75.8[e]
Last time	100.0	75.0	100.0	0.0	66.7	83.3	80.0	71.8	60.0	85.7	50.0	75.0	56.3	78.6	35.7	57.7

Note. Questions were worded as follows. *Had sex:* Have you ever done any of the following with someone? . . . Had sex? *Last 3 months:* (If yes) Have you had sex with someone in the last 3 months? *Next year:* (If no sex yet) Will you have sex with someone in the next year? *Used condoms:* Did you or your partner ever use condoms, which some people call rubbers? *Last time:* (If condom ever used) Did you use a condom the last time you had sex?

[a]Question not asked for these ages.
[b]Total based on ages 11–15 years (n = 98).
[c]Total based on ages 11–15 years (n = 34).
[d]Total based on ages 11–15 years (n = 95).
[e]Total based on ages 11–15 years (n = 66).

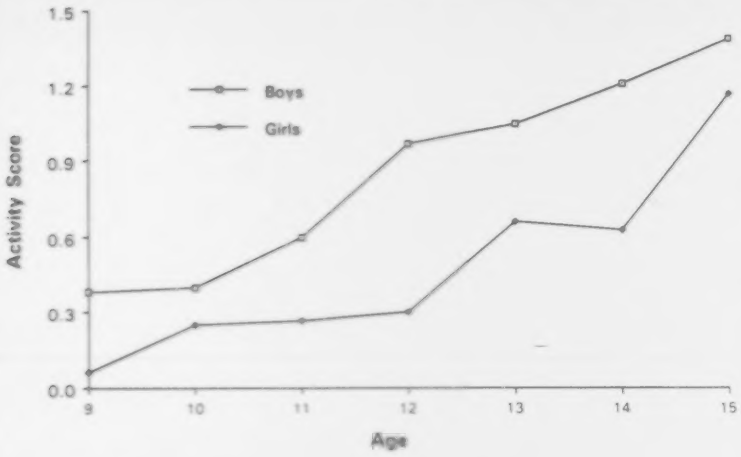


FIGURE 1—Mean sexual activity scores, by age and gender.

the results by age into two categories, 9 to 12 years of age and 13 to 15 years of age. A regression analysis of the condom use score (Table 3) revealed no significant differences by age or gender, although the tendency for girls to report more use of condoms was marginally significant ($P < .10$).

Table 4 contains the intercorrelation matrix for sexual activity and condom use with age, gender, and the other major variables. Because both age and gender were correlated with the other variables, we conducted multiple regression analyses to identify the unique relations of the other variables with sexual activity and condom use.

Parental Monitoring

Table 5 contains the results for the two parental monitoring scales. As seen in Table 4, girls perceived more of both types of parental monitoring, but only detailed knowledge was related to differences in age. Older children perceived that caretakers monitored the details of their social behavior less than younger children. To assess the relation between parental monitoring and sexual activity, we added the detailed knowledge scale as a predictor of sexual activity (Table 2). This analysis confirmed that greater monitoring was associated with less sexual activity holding constant the gender and age of children. There was no evidence of an interaction between age and parental monitoring, and monitoring was not associated with condom use (Tables 3 and 4).

Peer Behavior

Table 6 contains the results for perceived sexual experience and condom use of peers. As seen in Table 4, both scores were positively related to age, but the perceptions of boys did not differ from those of girls. Peer perceptions were also unrelated to parental monitoring. When perceived sexual experience of peers was used to predict sexual activity (Table 2), the interaction between age and perceived peer behavior was a significant predictor. As seen in Figure 2, the more that children perceived their peers as having engaged in sex, the greater their expression of sexual activity with age.

The regression analysis of condom use among sexually experienced children (Table 3) indicated that this protective behavior also varied as a function of the interaction between age and perceived peer behavior. As seen in Figure 3, children who perceived that most of their friends were using condoms were the only group to maintain condom use across age. Among children who perceived lower rates of condom use among friends, younger children tended to report more condom use than older children.

Attitude and AIDS Knowledge

To understand the influence of caretakers and peers, we explored the potential mediating roles of attitude toward sex and AIDS knowledge. Table 7 contains the results for the attitude toward sex and knowledge items. As seen in Table 4, attitude toward sex was directly related to age and was more favorable for boys than girls. As would be expected, attitude was also strongly related to sexual activity. To determine whether attitude could explain the associations of parental monitoring or peer perceptions with sexual activity, attitude was added to the multiple regression equations prior to the social variables. This analysis indicated that, with attitude in the equation, the impact of parental monitoring declined from $B = -.11$ ($P < .05$) to $B = -.09$ ($P < .10$). However, neither the simple relation with perceived peer behavior nor the interaction with age changed as a result of holding constant attitudes toward sex.

As a group, children tended to score above chance levels on each type of knowledge question (general misconceptions, condom protection, and sexual risks). This was especially true for questions about the risks of sexual transmission. However, children 9 years of age did not respond at a better than chance rate to the misconception and condom questions. Correct responses to these questions tended to increase with age and to increase more for girls than boys. None of the knowledge measures was related to sexual activity once age and gender were controlled; thus, these variables could not be used to explain the effects of parental monitoring or peer behavior. Condom use was also not associated with any of the knowledge measures, nor was it predicted by attitudes toward sex. We did not measure attitudes toward condoms per se.

Discussion

The rates of sexual activity we observed confirm our assumption that minority youths living in poverty represent a particularly high-risk population for exposure to HIV and other sexually transmitted diseases. Although our sample was young relative to national US samples, the present rate of sexual activity among 14- and 15-year-old boys was comparable to the rates reported for 18- and 19-year-old African-American boys in a representative survey conducted in 1988.[3] The present rate for girls at 15 years of age was comparable to the rates for girls 18 to 19 years of age in a different representative survey of African-American girls conducted in 1988.[2]

The present results also validate the hypothesis that differences in the onset and progression of sexual activity are related to social influences. The older the children in this sample, the more knowl-

TABLE 2—Hierarchical Regression Analysis of Sexual Activity Scale (Three Stages)

	Stage 1			Stage 2			Stage 3		
	B	*b*	95% CI	*B*	*b*	95% CI	*B*	*b*	95% CI
Age (linear)	.52*	.17*	.14, .20	.51*	.16*	.13, .19	.37*	.12*	.09, .15
Gender (boys = 1; girls = −1)	.31*	.19*	.13, .25	.29*	.18*	.12, .24	.26*	.16*	.11, .21
Gender × Age	.06	.02	−.01, .05	.04	.01	−.02, .04	.02	.01	−.02, .04
Detailed knowledge				−.11*	−.13*	−.25, −.01	−.07	−.09	−.20, .02
Detailed knowledge × Age				−.06	−.04	−.10, .02	−.03	−.02	−.08, .04
Peer sex perceptions (most = 1; some = 0; none = −1)							.36*	.26*	.19, .33
Peer sex perceptions × Age							.14*	.05*	.01, .09

Note. *B* is the standardized regression coefficient; *b* is the unstandardized coefficient; CI = confidence interval. R^2 change after each stage was .39 (stage 1), .01 (stage 2), and .09 (stage 3).
*$P < .05$.

TABLE 3—Hierarchical Regression Analysis of Condom Use Scale (Three Stages)

	Stage 1			Stage 2			Stage 3		
	B	*b*	95% CI	*B*	*b*	95% CI	*B*	*b*	95% CI
Age (9 to 12 = −1; 13 to 15 = 1)	.05	.01	−.19, .21	.06	.06	−.14, .26	.04	.04	−.16, .24
Gender (boys = 1; girls = −1)	−.19	−.14	−.30, .02	−.16	−.13	−.30, .04	−.16	−.13	−.29, .03
Gender × Age	−.12	−.11	−.31, .09	−.10	−.10	−.30, .10	−.09	−.09	−.28, .10
Detailed knowledge				.10	.17	−.16, .50	.11	.18	−.12, .48
Detailed knowledge × Age				.06	.05	−.12, .22	.04	.04	−.11, .19
Peer condom perceptions (most = 1; some, none = −1)							.42*	.35*	.19, .51
Peer condom perceptions × Age							.20*	.20*	.01, .39

Note. *B* is the standardized regression coefficient; *b* is the unstandardized coefficient; CI = confidence interval. R^2 change after each stage was .03 (stage 1), .01 (stage 2), and .16 (stage 3).
*$P < .05$.

edgeable they were about the hazards of sex and the benefits of condoms. In spite of this knowledge, their sexual activity increased as a function of age, especially when the perceived sexual behavior of peers was supportive. Importantly, they were more likely to engage in sex when they perceived low levels of monitoring by their caretakers and more likely to use condoms if they believed that their peers used them.

Although peer perceptions were more strongly related to sexual behavior than parental monitoring, it is not possible to draw conclusions about the relative magnitude of each influence. Because the test–retest reliabilities of parental monitoring scores were lower than those of peer perceptions, the resulting associations with sexual behavior may also have been lower. Nevertheless, the results do suggest that the two types of influence operated through different mechanisms. Monitoring was sensitive to both age and gender, as was attitude toward sex. Peer perception, however, was largely unrelated to monitoring, was sensitive only to age, and remained related to sexual activity once attitude and monitoring were held constant. This pattern suggests that monitoring was part of a process that influenced sexual attitudes and behavior separately from peer influence. It is not likely that greater monitoring reduced children's opportunities to engage in sex because this should have depressed the rate of change of sexual activity with age. However, there was no evidence of this effect.

The results for perceived peer behavior suggest that peers influenced normative boundaries for acceptable behavior rather than individual attitudes. If the motivation to engage in sex increases with age (as it appeared to do), then this increase should be expressed to a greater degree when peer norms are more favorable for sex. This would result in the observed greater rate of sexual activity with age for children with favorable group norms. It is also plausible that greater peer pressure for condom use would make nonuse of condoms less acceptable. If the individual motivation to use condoms declines as adolescents age (as considerable survey evidence suggests), then this decline should be more pronounced when peer norms are less favorable for condom use, precisely the result that we found.

The finding that condom use tended to decline with age except among children who perceived that most of their friends also used condoms suggests the powerful role that peer behavior can have on this important risk-reducing behavior. However, it is not clear why some children use condoms less as they mature. This repeated finding in surveys of adolescents requires further research.[5] One possibility is that, as children age, they find partners whom they regard as stable and feel less need for condoms. Another is that condom use interferes with sexual enjoyment and is difficult to maintain regardless of the partner. Finally, it is possible that

TABLE 4—Intercorrelations between Variables

	Sexual Activity	Condom Use	Age	Gender[a]	Monitoring: Detailed Knowledge	Monitoring: Whereabouts	Peer Sex Perception	Peer Condom Perception	Attitude Toward Sex	Knowledge: Misconceptions	Knowledge: Condoms
Condom use	.13										
Age	.54*	.01									
Gender[a]	.34*	−.13	.08								
Monitoring: detailed knowledge	−.23*	.11	−.16*	−.19*							
Monitoring: whereabouts	−.11	.12	.01	−.18*	.41*						
Peer sex perception	.50*	.08	.38*	.10	−.11	−.09					
Peer condom perception	.35*	.35*	.24*	.07	−.07	−.03	.45*				
Attitude toward sex	.50*	.01	.47*	.31*	−.20*	−.12	.34*	.20*			
Knowledge: misconceptions	.18*	−.09	.40*	−.03	−.09	.06	.18*	.06	.19*		
Knowledge: condoms	.16*	−.08	.28*	.01	−.06	−.02	.09	.15*	.15*	.33*	
Knowledge: sex risks	.13*	.02	.11	.08	.00	.07	.14*	.07	−.01	.08	.18*

Note. Most correlations were based on n = 300. Correlations with condom use were based on n = 116; correlations with peer perceptions of sexual experience were based on n = 282; correlations with peer perceptions of condom use were based on n = 276; and correlations with attitudes toward sex were based on n = 285.
[a]Boys = 1; girls = −1.
*$P < .05$.

TABLE 5—Perceptions of Parental Monitoring, by Age and Gender

	Age, y															
	Girls								Boys							
	9 (n = 32)	10 (n = 30)	11 (n = 24)	12 (n = 22)	13 (n = 25)	14 (n = 15)	15 (n = 12)	Total (n = 160)	9 (n = 20)	10 (n = 25)	11 (n = 24)	12 (n = 19)	13 (n = 21)	14 (n = 17)	15 (n = 14)	Total (n = 140)
Whereabouts scale	1.55	1.72	1.71	1.61	1.70	1.67	1.67	1.66	1.55	1.36	1.38	1.66	1.48	1.56	1.32	1.47
Detailed knowledge scale	1.61	1.57	1.54	1.56	1.33	1.63	1.40	1.53	1.45	1.44	1.29	1.45	1.24	1.38	1.07	1.34

Note. All scores are means ranging along a scale of 0 (never), 1 (sometimes), and 2 (always). Questions were worded as follows. *Whereabouts scale:* Do your parents know where you are when you are not at home or in school? Do your parents expect you to call them if you are going to be home late? *Detailed knowledge scale:* Do your parents want to know who you are going to be with before you go out? Do you talk with your parents about what you are going to do before you go out with friends? Do your parents ask you where you're going when you go out? Do you talk to your parents about the things you did when you've gone out with friends?

TABLE 6—Perceptions (%) of Peer Sexual Activity and Condom Use

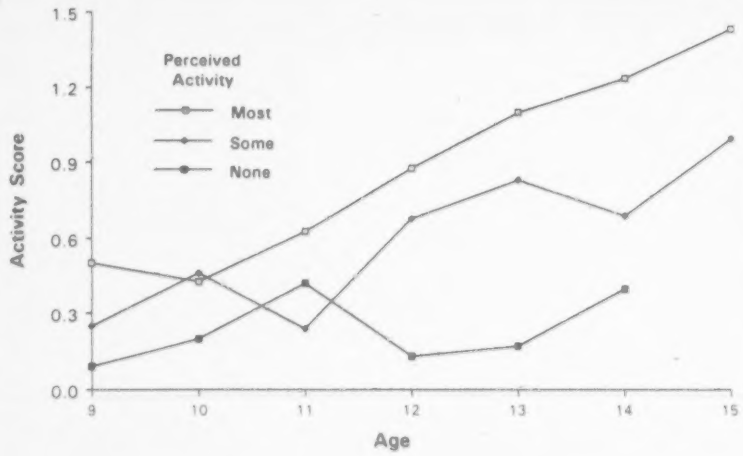
	Age, y															
	Girls								Boys							
	9 (n = 32)	10 (n = 30)	11 (n = 24)	12 (n = 22)	13 (n = 25)	14 (n = 15)	15 (n = 12)	Total (n = 160)	9 (n = 20)	10 (n = 25)	11 (n = 24)	12 (n = 19)	13 (n = 21)	14 (n = 17)	15 (n = 14)	Total (n = 140)
Friends had sex																
Most	15.6	23.3	29.3	36.4	40.0	40.0	50.0	30.6	15.0	32.0	37.5	42.1	47.6	64.7	78.6	42.9
Some	18.8	23.3	29.2	36.4	40.0	40.0	50.0	30.6	30.0	20.0	29.2	36.8	38.1	17.6	14.3	27.1
Friends use condoms (base = those who have had sex)																
Most	100.0	50.0	66.7	66.7	83.3	50.0	40.0	61.5	40.0	14.2	50.0	50.0	62.5	64.3	28.5	47.4
Some	0.0	25.0	0.0	33.3	0.0	50.0	50.0	25.6	20.0	14.2	0.0	33.3	12.5	7.1	28.5	16.7

Note. Questions were worded as follows. *Friends had sex:* Thinking about the boys and girls whose names you gave me earlier, would you say most of them, some of them, or none of them have had sex? *Friends use condoms:* (if "some" or "most" friends have had sex) For those who have had sex, do most of them, some of them, or none of them ever use a condom?

TABLE 7—Attitude toward Sex and AIDS Knowledge, by Age and Gender

	Age, y															
	Girls								Boys							
	9 (n = 32)	10 (n = 30)	11 (n = 24)	12 (n = 22)	13 (n = 25)	14 (n = 15)	15 (n = 12)	Total (n = 160)	9 (n = 20)	10 (n = 25)	11 (n = 24)	12 (n = 19)	13 (n = 21)	14 (n = 17)	15 (n = 14)	Total (n = 140)
Attitude toward sex[a]	1.43	1.90	1.96	3.00	3.04	2.60	2.92	2.30	2.05	2.52	2.76	3.53	3.75	4.00	4.07	3.15
Knowledge (mean no. correct)																
Misconceptions (4)[b]	1.75	2.63	2.54	3.27	3.16	3.53	3.75	2.78	1.95	2.40	2.70	2.95	2.86	3.12	3.29	2.71
Condom protection (2)[c]	1.00	1.23	1.33	1.64	1.68	1.80	1.75	1.42	1.25	1.56	1.46	1.37	1.14	1.65	1.71	1.44
Sexual risks (3)[d]	2.09	2.27	2.13	2.55	2.76	2.80	2.25	2.38	2.75	2.48	2.58	2.00	2.43	2.71	2.57	2.50

[a]Mean score on a scale of 1 to 5 based on response to "How would you feel if you were to have sex with someone in the next year? Would you feel very bad (1), bad (2), neither good nor bad (3), good (4), or very good (5) if you had sex in the next year?"
[b]Items used to measure misconceptions (the "correct" answer was "False" in each case): If you touch someone with AIDS, you can get AIDS. Only drug addicts get AIDS. You can tell if someone has AIDS just by looking at them. AIDS can be cured if treated early.
[c]Items used to measure condom protection (with "correct" answers): Using a condom during sex is a way to protect yourself from getting AIDS (True). Taking birth control pills is one way to protect yourself from becoming infected with the AIDS virus (False).
[d]Items used to measure sexual risks (the "correct" answer was "True" in each case): You can get AIDS the first time you have sex. You can get AIDS if you are only having sex with one person. It is important to use a condom even if you've been having sex with the same person for a while.



Note. The data point for 15-year-old children with no sexually active friends is omitted because only one child contributed to this mean.

FIGURE 2—Mean sexual activity scores, by age and proportion of friends perceived as sexually experienced (none, some, most).

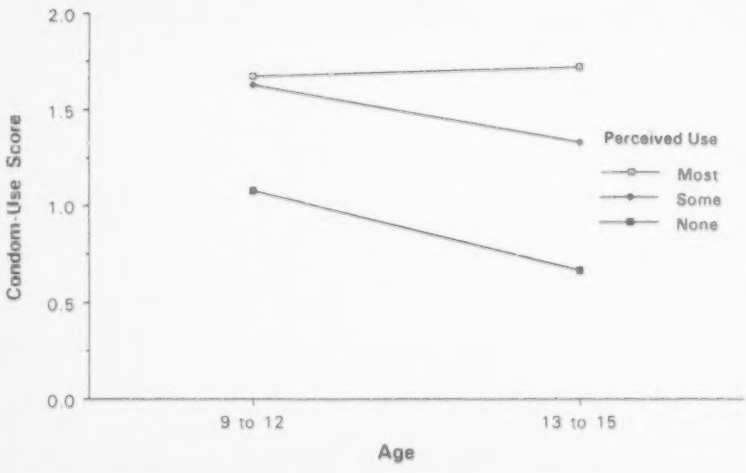


FIGURE 3—Mean condom use scores, by age and proportion of friends perceived to use condoms (none, some, most).

recent educational campaigns have succeeded in reaching younger children who are just becoming sexually active to a greater degree than older children. If this is true, our results could reflect a cohort effect rather than a stable trend that adolescents undergo as they become more sexually experienced.

Alternative Explanations

Previous research with adolescents[13,14] suggests that the development of sexual behavior in boys is largely dependent on hormonal levels and that these influences can explain sexual behavior better than social processes. However, consistent with more recent research,[15,24] the present results suggest that parental and peer behavior also exert influence on the sexual expression of boys and girls. To argue otherwise would require, in the case of monitoring, that parents monitor children less to the degree that children are sexually active, a possible although implausible explanation.

In the case of peer influence, a hormonal–maturational explanation might assume that children's perceptions of peers are simply a consequence of their own sexual behavior.[24] That is, children choose friends (or claim to associate with others) who are similar to themselves in interests and social characteristics, including sexual behavior. However, the data were inconsistent with this interpretation. If peer perception merely reflects individual behavior, the relation between age and sexual behavior should not depend on peer perceptions.

Another explanation consistent with similarity-choice mechanisms is that children's sexual practices become more important for friendship choices as children age, resulting in greater alignment between behavior and peer perception. If this interpretation were valid, however, it would only reinforce the importance of peer influences as a source of social support for risky sexual behavior and increase the need to consider social networks as a point of intervention.

Implications for Intervention

In this research, we have demonstrated that minority preadolescents and early adolescents living in poverty are at great risk for acquisition of HIV and other sexually transmitted infections. Through the use of sensitive interview techniques, we have demonstrated that these children's rates of reported sexual activity are comparable to those among young people 3 and 4 years older (according to nationally representative samples). These results indicate the need for interventions that target children in high-risk environments at earlier ages than national surveys would indicate.

This analysis also has several implications for intervention design. Consistent with the conclusions of many prevention researchers,[5] we found that knowledge about the hazards of unprotected sex is not sufficient to overcome the host of other influences that increase risky sexual behavior. Furthermore, our results suggest that even changing individual attitudes and skills may not undo the influence of peer norms. Even if parents could monitor their children more carefully, the rate at which their children's sexual activity progresses appears to be influenced by peers. Moreover, interventions designed to affect attitudes or abilities regarding condoms may have limited impact if peers do not support condom use. Nevertheless, our results suggest two strategies for reducing the prevalence of

risky sexual behavior in this age group. One strategy would be to encourage parents of children's friends to join together to monitor the behavior of their children. If parental monitoring were shared across friendship networks, peer influences might also transmit the influences of parental monitoring. Although shared caregiving is a traditional practice among African-American families,[25,26] parents' efforts may benefit from interventions designed to increase the shared monitoring of their children. Another strategy would be to promote the appeal of delayed sexual activity (or of continued use of condoms) to children within their friendship networks so that the social norms of peers would favor risk reduction rather than enhancement. □

Acknowledgments

This research was supported by grant U10MH48068 from the National Institute of Mental Health, AIDS-Prevention Studies Branch (Bonita Stanton, principal investigator).

We thank Dr Frank F. Furstenberg, Jr, and Dr Laurence Steinberg for their helpful comments on this paper.

References

1. Hein K. Adolescents at risk for HIV infection. In: DiClemente R, ed. *Adolescents and AIDS: A Generation at Risk.* Newbury Park, Calif: Sage; 1992:3–16.
2. Forrest JD, Singh S. The sexual and reproductive behavior of American women, 1982–1988. *Fam Plann Perspect.* 1990;22: 206–214.
3. Sonenstein FL, Pleck JH, Ku LC. Sexual activity, condom use and AIDS awareness among adolescent males. *Fam Plann Perspect.* 1989;21:152–158.
4. Centers for Disease Control. Selected behaviors that increase risk for HIV infection among high school students—United States, 1990. *MMWR.* 1992;41:231, 237–240.
5. DiClemente R. Psychosocial determinants of condom use among adolescents. In: DiClemente R, ed. *Adolescents and AIDS: A Generation at Risk.* Newbury Park, Calif: Sage; 1992:34–51.
6. Cates W. Acquired immunodeficiency syndrome, sexually transmitted diseases, and epidemiology: past lessons, present knowledge, and future opportunities. *Am J Epidemiol.* 1990;131:749–758.
7. Hein K. AIDS in adolescence. *J Adolesc Health Care.* 1989;10(suppl):10S–35S.
8. Fisher JD, Misovich SJ, Fisher WA. Impact of perceived social norms on adolescents' AIDS-risk behavior and prevention. In: DiClemente R, ed. *Adolescents and AIDS: A Generation at Risk.* Newbury Park, Calif: Sage; 1992:117–136.
9. Romer D, Hornik RC. HIV education for youth: the importance of social consensus in behavior change. *AIDS Care.* 1992;4:285–303.
10. Brooks-Gunn J, Petersen AC. Problems in studying and defining pubertal events. *J Youth Adolescence.* 1984;13:181–196.
11. Walter HJ, Vaughan RD, Galdis MM, Ragin DF, Kasen S, Cohall AT. Factors associated with AIDS risk behaviors among high school students in an AIDS epicenter. *Am J Public Health.* 1992;82:528–532.
12. Furstenberg FF Jr, Moore KA, Peterson JL. Sex education and sexual experience among adolescents. *Am J Public Health.* 1985;75:1331–1332.
13. Udry JR, Billy JOG, Morris NM, Groff TR, Raj MH. Serum androgenic hormones motivate sexual behavior in adolescent boys. *Fertil Sterility.* 1985;43:90–94.
14. Udry JR, Talbert EM, Morris NM. Biosocial foundations for adolescent female sexuality. *Demography.* 1986;23:217–227.
15. Brooks-Gunn J, Furstenberg FF Jr. Adolescent sexual behavior. *Am Psychol.* 1988;44: 249–257.
16. Baumrind D. Parenting styles and adolescent development. In: Brooks-Gunn J, Lerner R, Petersen AC, eds. *The Encyclopedia of Adolescence.* New York, NY: Garland; 1991:746–758.
17. Steinberg L, Dornbusch SM, Brown BB. Ethnic differences in adolescent development: an ecological perspective. *Am Psychol.* 1992;47:723–729.
18. Biglan A, Metzler CW, Wirt R, et al. Social and behavioral factors associated with high-risk sexual behavior among adolescents. *J Behav Med.* 1990;13:245–261.
19. Lamborn SD, Mounts NS, Steinberg L, Dornbusch SM. Patterns of competence and adjustment among adolescents from authoritative, authoritarian, indulgent, and neglectful families. *Child Dev.* 1991;62: 1049–1065.
20. Mays VM, Cochran SD. Issues in the perception of AIDS risk and risk reduction activities by black and Hispanic/Latina women. *Am Psychol.* 1988;43:949–957.
21. Stanton B, Aronson R, Borgatti S, Galbraith J, Feigelman S. Urban adolescent high-risk sexual behavior II: corroboration of focus group discussions through pile-sorting. *AIDS Educ Prev.* In press.
22. Kann L, Nelson GD, Jones JT, Kolbe LJ. Establishing a system of complementary school-based surveys to annually assess HIV-related knowledge, beliefs, and behaviors among adolescents. *J School Health.* 1989;59:55–58.
23. Kelly JA, St. Lawrence JS, Hood HV, Brasfield TL. An objective test of AIDS risk behavior knowledge: scale development, validation, and norms. *J Behav. Ther Exp Psychiatry.* 1989;20:227–234.
24. Udry JR. Biological predispositions and social control in adolescent sexual behavior. *Am Sociol Rev.* 1988;53:709–722.
25. Stack CB. *All Our Kin: Strategies for Survival in a Black Community.* New York, NY: Harper Torchbooks; 1974.
26. Taylor RL. Childrearing in African-American families. In: Everett JE, Chipungu SS, Leashore BR, eds. *Child Welfare: An Africentric Perspective.* New Brunswick, NJ: Rutgers University Press; 1991.

The Adequacy of Prenatal Care and Incidence of Low Birthweight among the Poor in Washington State and British Columbia

Steven J. Katz, MD, MPH, Robert W. Armstrong, MD, PhD, and James P. LoGerfo, MD, MPH

ABSTRACT

Objectives. The purpose of this study was to examine differences in adequacy of prenatal care and incidence of low birthweight between low-income women with Medicaid in Washington State and low-income women with Canadian provincial health insurance in British Columbia.

Methods. A population-based cross-sectional study was done by using linked birth certificates and claims data.

Results. Overall, the adjusted odds ratio for inadequate prenatal care in Washington (comparing women with Medicaid with those with private insurance) was 3.2. However, the risk varied by time of Medicaid enrollment relative to pregnancy (2.0, 1.0, 2.7, 6.3; for women who enrolled prior to pregnancy, during the first trimester, during the second trimester, or during the third trimester, respectively). In British Columbia, the adjusted odds ratio for inadequate care (comparing women receiving a health premium subsidy with those receiving no subsidy) was 1.5 for women receiving a 100% subsidy and 1.2 for women receiving a 95% subsidy. The risk for low birthweight followed a similar trend in both regions, but there was no association with enrollment period in Washington.

Conclusions. Overall, the risk for inadequate prenatal care among poor women was much greater in Washington than in British Columbia. Most of the difference was due to Washington women's delayed enrollment in Medicaid. In both regions, the poor were at similar risk for low birthweight relative to their more affluent counterparts. (*Am J Public Health.* 1994;84:986–991)

Introduction

New federal and state health policy initiatives may result in the gradual dissolution of Medicaid. In its place, federal and state governments may provide subsidies to the poor to sponsor insurance coverage under one of many regional health insurance purchasing groups. Thus, health insurance coverage for the poor may be merged with coverage for more affluent persons. These changes, already under way in some states such as Washington State, will have an important potential impact on health care for the poor.

A primary motivation for these reforms is the checkered performance of the Medicaid program. For instance, though Medicaid has been credited for narrowing the gap in infant morbidity and mortality between the poor and the more affluent, progress has stagnated.[1–3] Recent expansions of the program have shown little or no improvement in the utilization of prenatal care or outcomes among new Medicaid enrollees.[4–6]

Reasons for these failures fall into two broad categories.[7] First, despite Medicaid reforms, insurance-related barriers persist. Providers resist caring for patients whose insurance payments for prenatal care and delivery are much lower than prevailing community fees.[8,9] Because Medicaid eligibility may include the precondition of pregnancy, some women must first become pregnant before qualifying for coverage, which may result in late initiation of prenatal care.[10,11] Second, optimal insurance coverage alone does not address problems associated with poverty that may contribute to inadequate prenatal care and poor outcomes among the poor such as limited education, being unmarried, teenage pregnancy, and lack of transportation.[5,12] Although it is hoped that eliminating Medicaid will narrow the gap between the poor and more affluent, the results remain uncertain.

The Canadian health insurance program is an important benchmark for how proposed US reforms might affect health care for the poor because the program is uniform and universal. Thus, structural and economic barriers to care are minimized. Payments to physicians are uniform, which eliminates the effect of low payment on provider participation. Finally, the poor are not identified and potentially stigmatized by the insurance plans.

We compared differences in the adequacy of prenatal care and the incidence of low birthweight between poor and more affluent women in Washington State and British Columbia delivering during a period from 1987 through 1989. We chose these two regions because they are similar in demographic characteristics and in the organization of obstetrical care. We hypothesized that the elimination of barriers related to health insurance in Canada narrows the kind of gaps in adequacy of prenatal care and incidence of low birthweight between poor and

Steven J. Katz is with the Departments of Medicine and Health Services Management and Policy, University of Michigan, Ann Arbor. Robert W. Armstrong is with the Department of Pediatrics, University of British Columbia, Vancouver. James P. LoGerfo is with the Departments of Medicine and Health Services, University of Washington, Seattle.

Requests for reprints should be sent to Steven J. Katz, MD, MPH, Division of General Medicine, 3116 Taubman Center, University of Michigan Medical Center, Ann Arbor, MI 48109-0376.

This paper was accepted November 17, 1993.

more affluent women observed in the United States.

Methods

Study Population

The population was all singlet births in nonfederal hospitals in Washington State from July 1, 1988, through December 31, 1989 (n = 95 401) and in British Columbia from January 1, 1987, through December 31, 1988 (n = 79 638).

Data Source

In Washington, information from birth certificates, hospital discharge abstracts, and the Medicaid enrollment file was linked and combined for 96% of births (n = 95 401).[13] In British Columbia, information from birth certificates, physician claims, and hospital discharge abstracts was linked and combined for 92% of births occurring during the study period (n = 72 831). The distribution of demographic factors and health services utilization among women with linked records compared with women excluded because of unlinked records was similar in the two regions. Details of the record linkage process are available elsewhere.*

Definitions of Variables

Adequacy of prenatal care. The principal dependent variable was the modified Kessner index, a quantitative measure of the adequacy of care that considers the timing of the first prenatal visit (the month prenatal care began) and the total number of prenatal visits adjusted for gestational age.[14] A woman is categorized as receiving inadequate prenatal care if the first visit occurred during the third trimester of pregnancy or if the total number of visits was very low (regardless of the timing of care).

In Washington State, the variables for the number of prenatal visits, the month of pregnancy in which prenatal care began, and gestational age were obtained from the birth certificate. Gestational age was calculated from the date of the last menstrual period and the date of delivery. When only the day of the last menstrual period was missing (20% of cases), we imputed the missing day at 15. An additional variable, the clinical impression of gestational age, was used if the last-menses date was missing entirely. Eighty-six percent of records were complete for all variables in the analysis with the Kessner index.

Information about the month that prenatal care began may be more accurate than information on the number of visits.[15] The Washington State Department of Health validated these data from birth certificates by using medical records in 1987. Fifteen birth certificates from all delivery hospitals (n = 765) were selected. When present on the certificates, information regarding menses date, the month prenatal care began, and number of prenatal visits differed from medical records in 9.5%, 4.3%, and 9.7% of cases, respectively.[16] To minimize measurement error we eliminated prenatal visit utilization from the measure of adequacy of prenatal care by using only the month prenatal care began as the dependent variable. Inadequate care was defined as the initiation of care during the third trimester. Ninety-five percent of cases were complete for all variables when this variable was used in the analysis.

In British Columbia, information about use of prenatal services is not contained on birth certificates. Instead, we used claims data to calculate the timing of the first prenatal visit and the total number of prenatal visits. We obtained insurance enrollment dates to exclude women who did not continuously reside in the province during the study period. The month of pregnancy in which prenatal care began was calculated from the first prenatal visit date, the date of delivery, and the gestational age (obtained from birth certificates). Ninety-eight percent of records were complete for all variables.

Insurance status. In Washington State we used the Medicaid enrollment file to identify women whose delivery was insured under Medicaid. During the study period, any woman was eligible for Medicaid if the family income was 60% or less of the poverty level (approximately $7500 for a family of three). However, any pregnant woman could qualify if the family income was 100% or less of the poverty level (approximately $12 600 for a family of three). Self-pay and private insurance status was identified from the hospital payer code.

In British Columbia, all residents were required to pay a health insurance premium that was subsidized for low-income women. A resident received a 100% subsidy if he or she qualified for provincial income assistance or received a 95% subsidy if he or she had a taxable income that was low but not low enough to qualify for provincial income assistance. A family of three was eligible for the premium waiver if the annual gross income was under US$9600 or a 95% subsidy if the annual gross income was greater than US$9600 but less than US$16 000.

Demographic variables and birthweight. Birthweight, maternal age, marital status, and gravidity were obtained from birth certificates. Information on maternal race, obtained from birth certificates, was available only for Washington.

Analysis

We examined the adequacy of prenatal care and the incidence of low birthweight separately for each region using logistic regression. The first part of the analysis modeled the probability of inadequate prenatal care as a function of insurance status in Washington or subsidy status in British Columbia. The dependent variable was set equal to 1 if prenatal care was inadequate (Kessner score equal to 3) or 0 if prenatal care was adequate or intermediate (Kessner score equal to 1 or 2). A second dependent variable was modeled and set to 1 if the first prenatal visit occurred during the third trimester or 0 if the first prenatal visit occurred before the third trimester.

In the second part of the analysis we modeled the probability of low birthweight (less than 2500 g) as a function of insurance or subsidy status for each region separately. In Washington, the principal independent variable was insurance status: Medicaid (enrolled before pregnancy, during the first trimester, during the second trimester, or after the second trimester), private insurance, or self-pay. The reference category was private insurance. In British Columbia, the principal independent variable was subsidy status (complete waiver or 95% premium subsidy). The reference category was women who did not receive a subsidy.

In both parts of the analysis, we controlled for maternal age, number of

*See National Auxiliary Publication Service (NAPS) document 05097 for 3 pages of Appendix—The Adequacy of Prenatal Care and Incidence of Low Birthweight among the poor in Washington State and British Columbia: Data Sources and Database Construction. Please order directly from NAPS c/o Microfiche Publications, PO Box 3513, Grand Central Station, New York, NY 10163-3513. Enclose with your order (a) $7.75 for paper copy or (b) $4.00 for microfiche (US only funds from a US bank). For orders from outside the United States and Canada, add postage of $4.50 for paper copy or $1.50 for microfiche. There is a $15.00 invoicing charge for all orders filled before payment.

TABLE 1—Demographic Characteristics of Women Delivering in Washington State, 1988 through 1989

	Private Insurance (n = 63 629)	Medicaid Overall (n = 26 681)	Medicaid, by Period of Enrollment			
			Before Pregnancy (n = 8306)	During First Trimester (n = 5745)	During Second Trimester (n = 5079)	During Third Trimester (n = 7551)
Mean age, y	28.3	23.5	24.2	22.6	22.8	23.9
Age < 20 y, %	4.1	25.1	18.7	13.6	31.1	22.8
Single, %	9.9	56.4	65.3	63.1	58.9	39.8
Gravidity, mean no.	1.5	1.6	2.4	1.3	1.3	1.5
White, %	88.5	69.5	66.1	77.8	72.1	61.8
Black, %	1.9	7.0	10.6	5.6	6.0	4.1
Hispanic, %	3.1	12.6	7.3	7.1	10.1	23.4
Native American, %	1.4	5.8	8.1	4.5	4.8	4.5
Asian, %	5.0	5.0	8.0	5.0	7.0	6.2

Note. Women who were uninsured at time of delivery are excluded (n = 5091).

TABLE 2—Demographic Characteristics of Women Delivering in British Columbia, 1987 through 1988

	Received No Subsidy (n = 61 414)	Received 100% Subsidy (n = 4564)	Received 95% Subsidy (n = 6853)
Mean age, y	28.3	24.7	24.8
Age < 20 y, %	3.1	19.0	12.3
Single, %	12.3	58.0	43.5
Gravidity, mean no.	1.3	1.6	1.3

previous pregnancies, marital status, and race. We controlled for race by restricting a secondary analysis in Washington to non-Black and non-Hispanic women. This resulted in a racial composition in Washington that was comparable to that in British Columbia (predominantly Caucasian with similar proportions of Asian and Aboriginal minorities). The final logistic regression models contained covariates describing insurance status in Washington or subsidy status in British Columbia, maternal age, maternal age squared, dichotomous variables indicating the total number of previous pregnancies (0 to 8 or more), and a dichotomous variable indicating marital status. Second-order interaction terms were tested, but they were not included because they did not change the regression coefficients for insurance or subsidy status in the main models.

Results

Table 1 shows the distribution of maternal age, gravidity, marital status, and race by insurance status in Washington State. Demographic characteristics of women with Medicaid are broken down by period of enrollment. Overall, women with Medicaid were much more likely to be younger than 20 years, single, and non-White than were women with private insurance. Most women with Medicaid who enrolled before pregnancy qualified on the basis of Aid to Families with Dependent Children. These women were significantly poorer (based on eligibility requirements), older, more likely to be single, and more likely to have children than women enrolled during pregnancy. Importantly, 50% of Black women with Medicaid enrolled before pregnancy, whereas 70% of Hispanic women with Medicaid enrolled after the first trimester of pregnancy.

Table 2 shows the distribution of age, marital status, and gravidity by subsidy status in British Columbia. Women receiving no health premium subsidy were similar to women receiving private insurance in Washington. Women receiving a 100% subsidy were similar to women receiving Medicaid in Washington. The demographic factors of women receiving a 95% subsidy were intermediate between women receiving no subsidy and women receiving a 100% subsidy.

There was a wide distribution of Medicaid and subsidy deliveries across hospitals in both regions. In 75% of the 68 Washington hospitals delivering more than 100 women per year, the proportion of Medicaid births was 20% or more. In 75% of the 65 British Columbia hospitals with 100 or more deliveries per year, the proportion of subsidy births was 15% or greater. Similarly, there was a wide distribution of low-birthweight infants across hospitals in both regions.

Table 3 shows the number of observations and the proportion of women with inadequate prenatal care among women receiving Medicaid compared with women receiving private insurance in Washington and among women receiving a premium subsidy compared with women receiving no subsidy in British Columbia. Also shown are odds ratios and 95% confidence intervals with and without adjustment for covariates. The large differences between unadjusted and adjusted odds ratios reflect the distribution of age, marital status, and gravidity by payer status and the strong effect of these covariates on utilization. Women who enrolled before pregnancy (women receiving Aid to Families with Dependent Children) were at higher risk for inadequate prenatal care than were women receiving private insurance. Women who enrolled during the second trimester were at much higher risk and women who enrolled after the second trimester were at the highest risk.

Low-income women in British Columbia were also at risk for inadequate prenatal care and low-birthweight infants. The adjusted odds ratio for inadequate

prenatal care was somewhat lower for women receiving a 100% premium subsidy waiver than it was for women insured with Medicaid before pregnancy in Washington. Women receiving a 95% health premium subsidy had a lower risk of inadequate prenatal care than did women receiving a 100% subsidy. Importantly, these results were the same when the month prenatal care began was used as the dependent variable for inadequate prenatal care instead of the Kessner index.

Overall, women with Medicaid were at higher risk of having a low-birthweight infant than were women with private insurance in Washington (Table 4). The risk of having a low-birthweight infant was similarly high among women receiving a 100% subsidy compared with women receiving no subsidy in British Columbia. However, despite the strong association between enrollment and adequacy of prenatal care in Washington, there was no association between enrollment and the incidence of low birthweight. We repeated the analysis of inadequate prenatal care and low birthweight in Washington excluding Hispanic and Black women. Excluding these women had a negligible effect on the adjusted odds ratio for inadequate prenatal care among Medicaid recipients. However, it did result in a slight increase in the adjusted odds ratio for low birthweight among Medicaid recipients who enrolled during the third trimester (from 1.3 to 1.5). These results are likely explained by the risk of inadequate prenatal care and low birthweight among Hispanic women (predominantly of Mexican or Central American descent). Table 1 shows that these women represented only 4% of births, but they represented 13% of women with Medicaid and 23% of women with Medicaid who enrolled during the third trimester. However, Hispanic women who enrolled late in pregnancy were not at risk for low birthweight. These findings are consistent with results from other studies.[1,3,17] By contrast, Black women who enrolled late were at increased risk of having low-birthweight infants. However, Black women represented only 7% of women receiving Medicaid and only 4% of women with Medicaid who enrolled during the third trimester. Thus, overall, controlling for race by restriction had a minimal impact on the results.

TABLE 3—Odds Ratios for the Risk of Inadequate Prenatal Care for Women with Medicaid vs Private Insurance Coverage in Washington and for Premium Subsidy vs No Subsidy in British Columbia

Location/ Women's Insurance or Subsidy Status	No. Observations	Inadequate Prenatal Care, %	Odds Ratio (95% Confidence Interval)	
			Without Adjustments	With Adjustments[a]
Washington				
Private	54 200	2.3	Reference	Reference
Medicaid enrolled				
Before pregnancy	7 169	10.5	4.9 (4.5, 5.4)	2.0 (1.8, 2.3)
First trimester	5 024	5.0	2.2 (1.9, 2.5)	1.0 (0.9, 1.2)
Second trimester	4 473	11.4	5.4 (4.8, 6.0)	2.7 (2.4, 3.0)
Third trimester	6 499	19.1	9.8 (9.0, 10.7)	6.3 (5.7, 6.8)
Overall	23 165	12.1	5.8 (5.4, 6.2)	3.2 (2.9, 3.4)
British Columbia				
No subsidy	61 344	4.4	Reference	Reference
95% subsidy	6 853	7.7	1.7 (1.5, 1.9)	1.2 (1.1, 1.3)
100% subsidy	4 564	10.6	2.4 (2.2, 2.7)	1.5 (1.3, 1.7)

[a]Adjusted for age, marital status, and gravidity.

TABLE 4—Odds Ratios for the Risk of Low Birthweight for Women with Medicaid vs Private Insurance Coverage in Washington and Premium Subsidy vs No Subsidy in British Columbia

Location/ Women's Insurance or Subsidy Status	No. Observations	Low Birthweight, %	Odds Ratio (95% Confidence Interval)	
			Without Adjustments	With Adjustments[a]
Washington				
Private	62 654	3.6	Reference	Reference
Medicaid enrolled				
Before pregnancy	8 290	7.5	2.2 (2.0, 2.4)	1.6 (1.5, 1.8)
First trimester	5 741	6.0	1.7 (1.5, 1.9)	1.3 (1.2, 1.5)
Second trimester	5 074	6.3	1.8 (1.6, 2.0)	1.5 (1.3, 1.7)
Third trimester	7 544	6.3	1.5 (1.3, 1.7)	1.3 (1.2, 1.5)
Overall	26 649	6.3	1.8 (1.7, 2.0)	1.4 (1.3, 1.5)
British Columbia				
No subsidy	61 344	3.1	Reference	Reference
95% subsidy	6 853	3.9	1.2 (1.1, 1.4)	1.1 (0.9, 1.2)
100% subsidy	4 564	4.6	1.5 (1.3, 1.7)	1.3 (1.1, 1.6)

[a]Adjusted for age, marital status, and gravidity.

Discussion

We found that, compared with their more affluent counterparts, women receiving Medicaid in Washington State were much more likely to have inadequate prenatal care than were poor women in British Columbia (adjusted odds ratio of 3.2 in Washington compared with 1.5 in British Columbia). However, most of the difference between Washington and British Columbia was due to delayed enrollment in Medicaid. Women who enrolled in Medicaid before pregnancy were only slightly more likely to have inadequate care than were women in British Columbia (odds ratio of 2.0 in Washington vs 1.5 in British Columbia).

Although the period of Medicaid enrollment was strongly associated with the adequacy of prenatal care, it was not associated with the incidence of low birthweight. These results are consistent with other studies that have found little relationship between the utilization of prenatal care (as expressed by the timing and number of visits) and low birthweight.[1] Though many organizations have endorsed the importance of regular prenatal care visits in reducing the incidence of low birthweight, the relationship, overall, appears modest. The occurrence of low

birthweight is influenced by many complex biological and social factors that may not be greatly affected by prenatal care. Women who may particularly benefit from improved prenatal care because of being at high risk for adverse outcomes represent a small proportion of all women giving birth. Furthermore, utilization measures alone do not address the content of prenatal care visits, an important factor in the quality of care.[18] These factors limit the power of observational studies to detect associations between the utilization of prenatal care and birth outcomes. In Washington, this relationship may be further obscured by the racial composition of the study population. Hispanic women, who were not at increased risk for low-birthweight infants, constituted 23.4% of third-trimester enrollees in Washington Medicaid, but only 3.1% of those with private insurance.

Women who were enrolled in Medicaid before pregnancy represented a unique group among all Medicaid enrollees. These women, who qualified under federal (Aid to Families with Dependent Children) or state income assistance programs, were the poorest and most likely to be single with dependent children. Thus, they were most at risk for inadequate prenatal care and adverse outcomes. For instance, these women were at highest risk for low birthweight. However, even in this high-risk group the odds of inadequate prenatal care approached those for women receiving income assistance in British Columbia. This suggests that a substantial amount of the difference in the risk of inadequate prenatal care for this group may not be due to insurance-related barriers but rather to other social and structural factors related to poverty.

The primary limitation of this study is the use of different data sources for the determination of the adequacy of prenatal care in Washington and British Columbia. In Washington we used data obtained from patients or medical records and recorded on birth certificates, whereas in British Columbia, we used claims data to calculate the timing of the first prenatal visit and the total number of visits. Although claims data in Canada have been shown to be an accurate source of utilization data,[19] the linkage of all prenatal services may have been incomplete. Thus, recall bias in Washington and ascertainment bias in British Columbia limit the direct comparison of the proportions of inadequate prenatal care across regions. However, the differential effect of bias is minimized by comparing the relative differences in the adequacy of care between subgroups within each region.

The utilization of prenatal care may be affected by additional factors such as maternal education. Data on maternal education were not available from our data sources. However, data from other sources suggest that among women of childbearing age, the educational disparity between the poor and the more affluent is greater in the United States than in Canada.[20,21] If we extrapolate this observation to the maternal population, controlling for education in our analysis would likely further reduce differences in the risk of inadequate prenatal care among the poor between Washington State and British Columbia.

Differences in the regional organization of prenatal care services may also affect our results. The predominant difference between the two regions is the presence of managed care in Washington (12% of deliveries). However, health maintenance organization status was not a significant confounding factor in the relationship between Medicaid status and prenatal care use. Importantly, the distribution of deliveries by poor women and episodes of low birthweight across hospitals was similar in the two regions, indicating a similar level of regionalization of care.

Although the Medicaid program has clearly improved the process and outcome of prenatal care among the poor, the discouraging results of initial Medicaid expansion efforts have led to critical questions regarding the best strategies for making further gains. Current policy initiatives favor the gradual dissolution of the Medicaid program. Poor persons would be sponsored by government and combined with others into regional health insurance purchasing groups. With Canada as a benchmark, our findings suggest that such a strategy would greatly reduce the current disparity in the utilization of prenatal care between the poor and the more affluent. However, the gap in low birthweight appears to be resistant to the removal of insurance-related barriers alone. Clearly, targeted interventions that address social factors will be necessary to address this problem. Finally, the difference in the adequacy of prenatal care and birth outcomes between the poor and the more affluent in British Columbia may represent a minimum difference that could be expected if all economic and structural barriers related to health insurance were eliminated in the United States. □

Acknowledgments

The study was supported by grants from the Medical Education and Research Foundation of the American Medical Association and the Health of the Public Program of the PEW Charitable Trust and the John D. Rockefeller Foundation. We gratefully acknowledge the additional support from the British Columbia Ministry of Health Data Linkage Group, The Center for Health Services & Policy Research, the University of British Columbia, and the British Columbia Medical Services Foundation. Dr Katz was a Robert Wood Johnson Clinical Scholar during the time of this study.

This paper was presented in part at the 15th Annual National Meeting of the Society of General Internal Medicine, Washington, DC, May 1992.

We thank Vicky Hohner, Ron Armstrong, and Bo Green for data preparation and Fred Connell for useful comments on previous drafts of the manuscript.

References

1. Institute of Medicine. *Preventing Low Birth Weight.* Washington, DC: National Academy Press; 1985.
2. Oberg CN, Lia-Hoagberg B, Hodkinson E, Skovholt C, Vanman R. Prenatal care comparisons among privately insured, uninsured, and Medicaid-enrolled women. *Public Health Rep.* 1990;105:533–535.
3. Institute of Medicine. *Prenatal Care: Reaching Mothers, Reaching Infants.* Washington, DC: National Academy Press; 1988.
4. Piper JM, Ray WA, Griffin MR. Effects of Medicaid eligibility expansion on prenatal care and pregnancy outcome in Tennessee. *JAMA.* 1990;264:2219–2223.
5. Schwethelm B, Margolis LH, Miller C, Smith S. Risk status and pregnancy outcome among Medicaid recipients. *Am J Prev Med.* 1989;5:157–163.
6. Haas JS, Udvarhelyi SI, Morris CN, Epstein AM. The effect of providing health coverage to poor uninsured pregnant women in Massachusetts. *JAMA.* 1993;269: 87–91.
7. Schlesinger M, Kronebusch K. The failure of prenatal care policy for the poor. *Health Aff (Millwood).* 1990;9:91–111.
8. Perloff JD, Kletke PR, Neckerman KM. Physician's decisions to limit Medicaid participation: determinants and policy implications. *J Health Politics Law.* 1987;12: 221–234.
9. Committee on Health Care for Underserved Women. *Obstetric and Gynecologic Services for Indigent Women: Issues Raised by an ACOG Survey.* Washington, DC: American College of Obstetricians and Gynecologists; 1988.
10. Howell EM, Ellwood MR. Medicaid and pregnancy: issues in expanding eligibility. *Fam Plann Perspect.* 1991;23:123–128.
11. Howell EM, Brown GA. Prenatal, delivery, and infant care under Medicaid in three states. *Health Care Financing Rev.* 1989;10: 1–15.
12. Miller CL, Margolis LH, Schwethelm B,

Smith S. Barriers in implementation of prenatal care program for low income women. *Am J Public Health.* 1989;79:62–64.
13. Hohner VK, Armstrong RA. *Birth Event Records Project: Developing a Permanent Linked Research Database.* Olympia, Wash: Department of Health, Health Information Systems; 1990.
14. Kessner DM, ed. *Infant Death: An Analysis by Maternal Risk and Health Care.* Vol. 1. Washington, DC: National Academy Press; 1973.
15. Alexander GR, Tompkins ME, Peterson DJ, Weiss J. Source of bias in prenatal care utilization indices: implications for evaluating the Medicaid expansion. *Am J Public Health.* 1991;81:1013–1016.
16. David RJ. The quality and completeness of birth weight and gestational age data in computerized birth files. *Am J Public Health.* 1990;70:964–973.
17. Shiono PH, Kebanoff MA, Graubard MA, Berendes HW, Rhoads GG. Birth weight among women of different ethnic groups. *JAMA.* 1986;255:48–52.
18. Peoples-Sheps MD, Kalsbeek WD, Siegel E. Why we know so little about prenatal care nationwide: an assessment of required methodology. *Health Services Res.* 1988;23:359–380.
19. Roos LL, Nicol JP, Cageorge SM. Using administrative data for longitudinal research: comparisons with primary data collection. *J Chronic Dis.* 1987;40:41.
20. *Ontario Health Survey 1990.* Ontario, Canada: Ministry of Health; 1990.
21. *U.S. National Health Interview Survey 1990.* Hyattsville, Md: National Center for Health Statistics; 1990.

Call for Abstracts for the National Primary Care Nurse Practitioner Symposium

The 20th National Primary Care Nurse Practitioner Symposium will be held July 20–23, 1995, in Keystone, Colo, and is sponsored by the University of Colorado Health Sciences Center School of Nursing. The theme is "A Legacy of Leadership: NPs Defining the Future."

Abstracts of research, innovative clinical practice/models and roles, and other professional papers relating to primary health care nursing are solicited. Preference will be given to abstracts with strong clinical applicability, innovative role development/practice models, and papers with a future-oriented perspective. Abstracts are being accepted for oral or poster presentation.

The guidelines for submission are the following:

- Send three typed copies of the abstract, limited to one page (300 words or less), single spaced, with 1½" left and 1" right margins. One copy, which will be printed in the syllabus if accepted, should include author name and contact information. The other two should have no author identification. Abstracts must include title and be photocopy-ready (do not use a dot matrix printer).
- Include author vita with author contact information.
- Format includes title, brief statement of purpose, methods, conclusions, and clinical application(s) or, if not a research abstract, brief summary of project/paper.
- Indicate interest in oral presentation and/or poster session.
- Presenters are provided tuition waiver for registration the day of their presentation, but they are responsible for their own travel and expenses.

Submission deadline is *September 15, 1994.* Contact Ellen Lemberg, RNC, MS, Nurse Practitioner Symposium Office, University of Colorado Health Sciences Center School of Nursing, 4200 E 9th Ave, Box C287, Denver, CO 80262; tel (303) 270-7436; fax (303) 270-3785.

Public Health Briefs

Smoking and Mortality among Residents of a California Retirement Community

ABSTRACT

Smoking cessation decreases mortality among the elderly. Participants in the Leisure World Cohort Study initiated in 1981 were followed until death or January 1, 1991. The 8869 women and 4999 men (median age = 73 years at initial survey) contributed 105 952 person-years of follow-up; 4002 had died. All-cause mortality rates were highest among current smokers; compared with never smokers the age-adjusted relative risks (and 95% confidence intervals) were 1.67 (1.46, 1.92) for women and 1.95 (1.66, 2.30) for men. Current smokers had increased risks of coronary heart disease, other cardiovascular disease, and cancer. Risk of death from smoking-related cancers continued to be high among former smokers, although lower than the risk among current smokers. Relative risks of mortality from cancer and cardiovascular disease increased with the number of cigarettes smoked per day and with a decreasing number of years since cessation of smoking. (*Am J Public Health.* 1994;84:992–995)

Annlia Paganini-Hill, PhD, and Grace Hsu, MS

Introduction

Although cigarette smoking causes coronary heart disease and cancer in middle age, few studies have examined the effect of continued smoking among those who survive to old age.[1-3] LaCroix et al. found that smokers' mortality rates were twice those of never smokers in a prospective study of 7178 persons aged 65 years or older.[1] Current smokers had higher rates of both cardiovascular and cancer mortality. Similarly, the Honolulu Heart Study found that smokers continued to have an increased risk of heart disease in older age.[2] However, the Framingham Heart Study reported no relation between smoking and mortality after 30 years of follow-up.[3] We evaluated the effect of smoking in a population-based cohort comprising a large number of people older than 75 years (including many women) and followed prospectively for 9½ years.

Methods

In June 1981, a health survey was mailed to all residents of Leisure World, Laguna Hills, a retirement community near Los Angeles, Calif. Residents moving into the community after this date were mailed this questionnaire in June 1982, June 1983, and October 1985. Residents are primarily White, affluent, well educated, and female (two thirds). The residents' median age at initial survey was 73 years. Of the 22 781 residents, 13 981 (61%), including 8879 women, returned questionnaires. Details of the study have been reported previously.[4,5]

The health survey requested information on life-style practices including use of cigarettes. Smoking status was classified as *current, past,* or *never* based on response to the question, "Have you *ever* smoked cigarettes during any period of your life (aside from possibly trying them once or twice)?" Current and past smokers were asked, "What is the greatest number of cigarettes that you have regularly smoked?" with the choices of *¼, ½, 1, 1½,* and *2 or more packs per day.* Current and past smokers were also asked, "At what age did you start smoking regularly?" Past smokers were asked, "At what age did you last stop?"

Cohort members were followed until death or January 1, 1991. Twenty cohort members were lost to follow-up; a search of the National Death Index did not reveal that these individuals were deceased. Cause of death as recorded on the death certificates was coded by using the *International Classification of Diseases,* Ninth Revision (ICD-9).

We evaluated the relation of smoking to mortality from all causes and from selected specific causes: coronary heart disease (ICD-9 codes 410–414); other cardiovascular disease (codes 390–405, 415–459, and 798); total cardiovascular disease (codes 390–459 and 798); smoking-related cancers (codes 141–145, 150, 157, 161, 162, 188, and 189), that is, cancers of the lung, trachea, bronchus, oral cavity, larynx, esophagus, pancreas, bladder, and kidney; and all cancers (codes 140–208).

Age-adjusted relative risks, with five age groups (≤ 74, 75–79, 80–84, 85–89, and 90+ years), and two-sided *P* values

The authors are with the Department of Preventive Medicine, University of Southern California School of Medicine, Los Angeles, Calif.

Requests for reprints should be sent to Annlia Paganini-Hill, PhD, Preventive Medicine, University of Southern California School of Medicine, 1721 Griffin Ave #103, Los Angeles, CA 90031.

This paper was accepted September 30, 1993.

were obtained by using a regression method that assumed that the occurrence of death could be regarded as a Poisson process with a constant hazard rate for a given person.[6,7] The GLIM statistical software package program was used to make these calculations.[8]

Results

The 8869 women and 4999 men providing information on smoking contributed 105 952 person-years of follow-up (70 309 for women and 35 643 for men). Deaths occurred in 1987 women and 2015 men. More women than men were current smokers at baseline (12% vs 9%), and the prevalence of smoking decreased with age in both sexes (Table 1). Age-adjusted relative risks of specific causes of death by various smoking categories are shown in Tables 2 and 3 for women and men, respectively.

Mortality rates from all causes were nearly double among current smokers compared with never smokers. The age-adjusted relative risk of total mortality was 1.67 (95% confidence interval [CI] = 1.46, 1.92) in women and 1.95 (CI = 1.66, 2.30) in men. The relative risks for current smokers were increased in each age group (≤74, 75–79, 80–84, 85–89, and 90+ years) except for men aged 90 years or older (2.08, 1.83, 1.55, 1.26, and 1.59 for women and 2.09, 1.45, 2.09, 1.63, and 0.50 for men). Both female and male former smokers had rates of total mortality between those of current smokers and never smokers; the age-adjusted relative risk was 1.20 (CI = 1.09, 1.33) for women and 1.24 (CI = 1.13, 1.37) for men.

The higher risk of death among current smokers was because of deaths from cardiovascular disease and cancer.

TABLE 1—Distribution of Smoking Status at Initial Survey, by Sex and Age: The Leisure World Cohort

	Females (n = 8869), %			Males (n = 5102), %		
Age, y	Current Smoker (n = 1106)	Past Smoker (n = 2880)	Never Smoker (n = 4883)	Current Smoker (n = 438)	Past Smoker (n = 2953)	Never Smoker (n = 1708)
<65	21	36	43	16	50	34
65–69	18	39	43	12	55	33
70–74	14	37	49	10	58	31
75–79	9	30	60	7	60	33
≥80	4	21	74	3	60	37

TABLE 2—Age-Adjusted Relative Risks (RRs) for Cause-Specific Mortality, by Smoking Categories: Females

		Cause of Death and ICD-9 Code											
		All Causes 1–999		CHD 410–414		Other CVD 390–405, 415–459, 798		Total CVD 390–459, 798		Smoking Cancers 141–145, 150, 157 161–162, 188–189		All Cancers 140–208	
Smoking Category	Person-Years[a]	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Smoking													
Never[b]	38 761	1.00		1.00		1.00		1.00		1.00		1.00	
Past	22 775	1.20	1.09, 1.33	1.28	1.06, 1.55	1.18	0.97, 1.45	1.24	1.08, 1.42	1.51	1.01, 2.25	1.00	0.81, 1.23
Current	8 725	1.67	1.46, 1.92	1.47	1.10, 1.95	1.93	1.48, 2.53	1.68	1.38, 2.05	3.70	2.41, 5.67	1.65	1.27, 2.14
Test for trend[c]		$P < .001$		$P < .01$		$P < .001$		$P < .001$		$P < .001$			
No. of years smoked													
1–19	6 336	1.06	0.89, 1.26	1.06	0.76, 1.47	1.22	0.88, 1.69	1.14	0.90, 1.43	0.85	0.39, 1.87	0.89	0.62, 1.28
20–39	12 114	1.10	0.96, 1.26	1.31	1.02, 1.67	1.12	0.85, 1.47	1.22	1.10, 1.46	1.38	0.83, 2.30	0.90	0.68, 1.19
≥40	11 693	1.68	1.49, 1.89	1.59	1.27, 2.00	1.51	1.18, 1.92	1.55	1.31, 1.83	3.59	2.45, 5.28	1.62	1.29, 2.03
Test for trend		$P < .001$		$P < .001$		$P < .01$		$P < .001$					
No. of packs/day													
¼–½	15 987	1.17	1.04, 1.31	1.15	0.93, 1.43	1.25	1.00, 1.55	1.20	1.03, 1.40	1.33	0.84, 2.10	1.03	0.81, 1.30
≥1	14 199	1.58	1.40, 1.78	1.65	1.32, 2.07	1.54	1.21, 1.96	1.60	1.35, 1.89	3.04	2.05, 4.51	1.38	1.10, 1.74
Test for trend		$P < .001$		$P < .001$		$P < .001$		$P < .001$		$P < .001$		$P < .05$	
Years since last smoked													
>20	8 575	0.99	0.85, 1.15	1.09	0.83, 1.42	1.03	0.78, 1.36	1.06	0.87, 1.28	0.86	0.44, 1.69	0.87	0.64, 1.19
11–20	6 576	1.32	1.13, 1.55	1.53	1.14, 2.04	1.13	0.80, 1.60	1.34	1.08, 1.67	1.60	0.87, 2.94	1.12	0.81, 1.55
6–10	3 222	1.32	1.05, 1.65	1.42	0.93, 2.15	1.26	0.79, 2.00	1.34	0.98, 1.83	2.05	0.97, 4.31	1.02	0.64, 1.63
≤5	3 391	1.52	1.22, 1.88	1.28	0.81, 2.01	1.45	0.92, 2.28	1.36	0.98, 1.87	3.30	1.79, 6.11	1.12	0.71, 1.75
Test for trend		$P < .001$		$P < .01$		$P < .001$		$P < .001$		$P < .001$		$P < .001$	
Age quit smoking													
<65	17 312	1.08	0.96, 1.22	1.17	0.93, 1.46	1.07	0.84, 1.37	1.12	0.95, 1.32	1.30	0.82, 2.07	0.97	0.76, 1.23
≥65	4 575	1.45	1.25, 1.69	1.55	1.18, 2.04	1.25	0.92, 1.71	1.41	1.14, 1.73	2.36	1.36, 4.12	1.11	0.78, 1.58
Test for trend		$P < .001$		$P < .001$		$P < .001$		$P < .001$		$P < .001$		$P < .001$	

Note. CHD = coronary heart disease; CVD = cardiovascular disease; CI = confidence interval.
[a]Person-years do not always total 70 309 because of missing values for some persons.
[b]Reference category for all relative risks.
[c]*P* values for test for linear trend are given unless test for deviation from linearity was significant ($P < .05$).

TABLE 3—Age-Adjusted Relative Risks (RRs) for Cause-Specific Mortality, by Smoking Categories: Males

Smoking Category	Person-Years[a]	Cause of Death and ICD-9 Code: All Causes 1–999 RR	95% CI	CHD 410–414 RR	95% CI	Other CVD 390–405, 415–459, 798 RR	95% CI	Total CVD 390–459, 798 RR	95% CI	Smoking Cancers 141–145, 150, 157, 161–162, 188–189 RR	95% CI	All Cancers 140–208 RR	95% CI
Smoking													
Never[b]	12 325	1.00		1.00		1.00		1.00		1.00		1.00	
Past	20 408	1.24	1.13, 1.37	1.16	0.98, 1.37	1.01	0.83, 1.24	1.10	0.97, 1.25	2.45	1.62, 3.71	1.42	1.14, 1.75
Current	2 907	1.95	1.66, 2.30	1.41	1.03, 1.93	1.91	1.36, 2.70	1.60	1.27, 2.02	6.39	3.83, 10.6	2.74	2.01, 3.74
Test for trend[c]				$P < .05$						$P < .001$		$P < .001$	
No. of years smoked													
1–19	4 560	0.93	0.79, 1.09	0.75	0.56, 1.00	0.88	0.63, 1.23	0.80	0.64, 1.00	1.01	0.49, 2.07	1.14	0.82, 1.59
20–39	9 582	1.16	1.03, 1.31	1.13	0.92, 1.38	1.02	0.80, 1.31	1.09	0.93, 1.27	1.86	1.14, 3.04	1.28	0.99, 1.66
≥40	8 400	1.68	1.51, 1.88	1.48	1.22, 1.79	1.33	1.05, 1.68	1.42	1.25, 1.64	4.85	3.18, 7.42	2.08	1.65, 2.63
Test for trend						$P < .05$						$P < .001$	
No. of packs/day													
¼–½	6 633	1.16	1.03, 1.32	1.19	0.96, 1.47	0.99	0.76, 1.29	1.11	0.94, 1.30	1.26	0.71, 2.24	1.21	0.92, 1.60
≥1	15 731	1.41	1.27, 1.56	1.21	1.01, 1.44	1.17	0.94, 1.46	1.19	1.04, 1.37	3.65	2.41, 5.53	1.72	1.38, 2.14
Test for trend		$P < .001$		$P < .05$		$P < .10$		$P < .05$		$P < .001$		$P < .001$	
Years since last smoked													
>20	10 225	1.09	0.98, 1.22	1.03	0.85, 1.25	0.95	0.75, 1.20	1.00	0.86, 1.16	1.77	1.10, 2.84	1.31	1.03, 1.67
11–20	5 732	1.36	1.19, 1.55	1.21	0.96, 1.53	1.21	0.91, 1.60	1.21	1.01, 1.45	2.91	1.76, 4.80	1.44	1.08, 1.92
6–10	2 125	1.48	1.22, 1.80	1.43	1.02, 1.99	0.86	0.52, 1.42	1.20	0.91, 1.58	3.91	2.11, 7.25	1.72	1.16, 2.55
≤5	1 670	2.00	1.64, 2.45	2.02	1.44, 2.84	1.38	0.84, 2.25	1.77	1.34, 2.34	4.10	2.07, 8.12	1.92	1.24, 2.96
Test for trend		$P < .001$		$P < .001$		$P < .001$		$P < .001$		$P < .001$		$P < .001$	
Age quit smoking													
<65	16 425	1.11	1.00, 1.24	1.03	0.86, 1.23	0.98	0.79, 1.22	1.01	0.88, 1.16	1.99	1.28, 3.08	1.30	1.04, 1.63
≥65	3 363	1.74	1.52, 1.98	1.71	1.37, 2.14	1.19	0.89, 1.60	1.49	1.25, 1.78	4.22	2.58, 6.91	1.92	1.44, 2.56
Test for trend						$P < .01$				$P < .001$		$P < .001$	

Note. CHD = coronary heart disease; CVD = cardiovascular disease; CI = confidence interval.
[a]Person-years do not always total 35 643 because of missing values for some persons.
[b]Reference category for all relative risks.
[c]*P* values for test for linear trend are given unless test for deviation from linearity was significant ($P < .05$).

The age-adjusted relative risks for all cardiovascular disease for current smokers were 60% greater than those of never smokers: 1.68 (CI = 1.38, 2.05) for women and 1.60 (CI = 1.27, 2.02) for men. This increased risk was observed for mortality from coronary heart disease and from other cardiovascular disease. No elevation in the rate of all cardiovascular disease was observed among male former smokers, although the rate in female former smokers was increased.

The rates of death from all cancers were higher among the current smokers than among the never smokers. The age-adjusted relative risk was 1.65 (CI = 1.27, 2.14) for women and 2.74 (CI = 2.01, 3.74) for men. Current smokers had especially high rates of mortality from smoking-related cancers compared with never smokers. The age-adjusted relative risk was 3.70 (CI = 2.41, 5.67) for women and 6.39 (CI = 3.83, 10.6) for men. The risk of death from smoking-related cancers continued to be high among former smokers, although it was lower than the risk among current smokers. The age-adjusted relative risks for former smokers were 1.51 (CI = 1.01, 2.25) and 2.45 (CI = 1.62, 3.71) for women and men, respectively.

Risks of death from cancer and cardiovascular disease increased with the number of cigarettes smoked per day in both sexes. Less consistent trends were observed for the number of years of smoking. Former smokers who quit smoking more than 20 years ago had rates of death similar to those among never smokers for each specific cause except cancer in men, the risk for which remained high. In all cases the risk of death increased with decreasing number of years since cessation of smoking. Likewise, for all causes of death and each cause-specific category, the relative risk of death was higher among those who quit smoking at age 65 years or later than it was among those who quit before age 65.

Discussion

Previous reports concerning middle-aged populations have linked current smoking to mortality from coronary disease and cancer. However, few studies have evaluated the association between smoking and mortality in people older than 65 years, and only one has had a substantial number of subjects older than 75 years of age. Our findings suggest that the increased risks associated with smoking extend to groups beyond age 75. This agrees with the recent report by LaCroix et al.,[1] who found a marginally significant increased risk of death among smokers in that age group.

Previously, we found little evidence of response bias in this cohort when comparing early and late respondents on most variables.[9] In terms of illness (hypertension, angina, myocardial infarction, stroke, diabetes, rheumatoid arthritis, glaucoma, and cancer), we found no large differences between early and late respon-

dents. As with other surveys, we found that current smokers were slower to respond and former smokers were faster to respond than never smokers. However, the differences were not large. Nonetheless, systematic differences between early and late responders and the slightly higher incidence of hospitalizations during the first year of study among nonrespondents indicate that the early respondents are the "worried well," i.e., healthy individuals who see their doctor regularly, receive disease detection screening, and follow healthful life-style practices. However, using internal standards for comparing mortality rates, as we did, is unlikely to produce biased results.

The findings of this study support the efforts toward smoking cessation. Cancer and cardiovascular disease are essentially diseases of older persons, and the positive effects of smoking cessation on these diseases extend to old age. In view of the finding that former smokers who quit less than 20 years ago remain at increased risk of death, smoking cessation should begin early to reduce the risks of cardiovascular and cancer mortality and increase the health of older adults. □

Acknowledgments

This work was supported by grant CA32197 from the National Cancer Institute, National Institutes of Health, Bethesda, Md.

We gratefully acknowledge the invaluable assistance of our research staff—Beverly Ducey and Marian Hawk—and are indebted to the residents of Leisure World, Laguna Hills, Calif, whose cooperation made this work possible.

References

1. LaCroix AZ, Lang J, Scherr P, et al. Smoking and mortality among older men and women in three communities. *N Engl J Med.* 1991;324:1619–1625.
2. Benfante RJ, Reed DM, MacLean C, Yano K. Risk factors in middle age that predict early and late onset of coronary heart disease. *J Clin Epidemiol.* 1989;42:95–104.
3. Harris T, Cook EF, Kannel WB, Coldman L. Proportional hazards analysis of risk factors for coronary heart disease in individuals aged 65 or older: the Framingham Heart Study. *J Am Geriatr Soc.* 1988;36:1023–1028.
4. Paganini-Hill A, Ross RK, Henderson BE. Prevalence of chronic disease and health care practices in a retirement community. *J Chronic Dis.* 1986;39:699–707.
5. Paganini-Hill A, Chao A, Ross RK, Henderson BE. Aspirin use and chronic diseases: a cohort study of the elderly. *BMJ.* 1989;299: 1247–1250.
6. Breslow NE, Lubin JH, Marek P, Langholz B. Multiplicative models and cohort analysis. *J Am Stat Assoc.* 1983;78:1–12.
7. Fromme EL. The analysis of rates using Poisson regression models. *Biometrics.* 1983; 39:665.
8. Royal Statistical Society. *Generalized Linear Interactive Modeling. The GLIM System, Release 3.* Oxford, England: Numerical Algorithms Group; 1978.
9. Paganini-Hill A, Hsu G, Chao A, Ross RK. Comparison of early and late respondents to a postal health survey questionnaire. *Epidemiology.* 1993;4:375–379.

The Utility of HMO Data for the Surveillance of Chronic Diseases

David L. Nordstrom, MS, MPH, Patrick L. Remington, MD, MPH, and Peter M. Layde, MD, MSc

ABSTRACT

In an area served by a single medical center that operates as both health maintenance organization (HMO) and fee-for-service clinic, we reviewed existing computerized medical records to determine the prevalence of 11 diseases. Standardized medical care utilization prevalence ratios, comparing the annual prevalences in the two groups, varied from 1.38 for rheumatoid arthritis to 0.60 for liver cirrhosis. Unless supplemented by data from hospitals, physicians, and other sources, HMO data may result in invalid estimates of the prevalence of chronic disease. (*Am J Public Health.* 1994;84:995–997)

Introduction

Surveillance of disease is the continuing scrutiny of all aspects of occurrence and spread of a disease that are pertinent to effective control.[1] Because in the United States chronic diseases account for a majority of deaths,[2] the Centers for Disease Control and Prevention are evaluating improved methods of surveillance.[3]

Traditionally, surveillance of chronic diseases has been done by using mortality data.[4] Although some registries track nonfatal cancer, cardiovascular disease, stroke, birth defects, trauma, and other noninfectious conditions,[5] few have employed routinely collected morbidity data from hospital discharge, health maintenance organization (HMO), or ambulatory care records.

Several characteristics of HMOs make them attractive for research, especially the existence of an identified and relatively stable population.[6,7] Yet examples of the use of clinical databases to relate the experience of the HMO population to epidemiologic trends are limited.[8,9] Studies of HMO populations are relatively recent.[10] No studies address the validity of such research, particularly with respect to disease surveillance.

Methods

Study Population

Marshfield Clinic is a 400-physician group practice adjacent to Saint Joseph's Hospital in Marshfield, Wis. The clinic

David L. Nordstrom and Peter M. Layde are with the Marshfield Medical Research Foundation, Marshfield, Wis. Patrick L. Remington is with the Bureau of Public Health, Wisconsin Department of Health and Social Services, Madison, Wis.

Requests for reprints should be sent to David L. Nordstrom, MS, MPH, Marshfield Medical Research Foundation, 1000 N Oak Ave, Marshfield, WI 54449-5790.

This paper was accepted September 1, 1993.

TABLE 1—Standardized Prevalence of Medical Care Utilization, by Disease and HMO Status, Marshfield Epidemiologic Study Area, 1986 through 1988

Disease	Rate per 10 000 Person-Years (No. of Cases)	
	HMO	Non-HMO
Rheumatoid arthritis	40.1 (519)	29.1 (185)
Breast cancer	67.9 (490)	53.2 (166)
Myocardial infarction	201.9 (2917)	165.4 (1051)
Colorectal cancer	17.1 (252)	15.0 (95)
Stroke	62.3 (922)	55.2 (351)
Diabetes	197.2 (2604)	188.1 (1195)
Chronic lung disease	117.1 (1650)	115.1 (731)
Epilepsy	63.9 (527)	76.8 (488)
Cervical cancer	8.7 (39)	11.5 (36)
Lung cancer	8.1 (118)	11.0 (70)
Cirrhosis of liver	3.4 (49)	5.7 (36)

Note. Listed in descending order of prevalence ratios, as shown in Table 2.

TABLE 2—Standardized Prevalence Ratios of Medical Care Utilization for Selected Diseases: HMO Members Compared with Non-HMO Members, by Sex, Marshfield Epidemiologic Study Area, 1986 through 1988

Disease	Males		Females		Total	
	SPR	95% CI	SPR	95% CI	SPR	95% CI
Rheumatoid arthritis	1.73	1.36, 2.20	1.25	1.02, 1.54	1.38	1.16, 1.64
Breast cancer	. . .	. . .	1.28	1.06, 1.53	. . .	. . .
Myocardial infarction	1.46	1.31, 1.61	0.95	0.85, 1.05	1.22	1.14, 1.31
Colorectal cancer	1.16	0.82, 1.63	1.13	0.81, 1.59	1.15	0.90, 1.46
Stroke	1.18	0.98, 1.41	1.08	0.90, 1.28	1.13	1.00, 1.28
Diabetes	1.34	1.22, 1.48	0.84	0.76, 0.92	1.05	0.98, 1.13
Chronic lung disease	1.10	0.96, 1.26	0.88	0.75, 1.02	1.02	0.93, 1.11
Epilepsy	0.84	0.70, 1.00	0.83	0.69, 0.99	0.83	0.73, 0.95
Cervical cancer	. . .	. . .	0.76	0.47, 1.22	. . .	. . .
Lung cancer	0.67	0.41, 1.11	0.89	0.52, 1.51	0.74	0.54, 1.00
Cirrhosis of liver	0.57	0.30, 1.10	0.64	0.33, 1.22	0.60	0.38, 0.95

Note. SPR = age-standardized prevalence ratio; CI = confidence interval. Ratios are listed in descending order.

and hospital maintain a combined paper and computerized medical record on all patients.

The Marshfield Epidemiologic Study Area comprises 12 postal (zip code) areas joined for epidemiologic, clinical, and health services studies.[11] Apart from race and ethnicity, population characteristics of the study area and the United States are similar.[12]

Virtually all primary, secondary, and tertiary medical care to residents of this nonmetropolitan area is provided by Marshfield Clinic. An estimated 79% of residents sought care at Marshfield Clinic at least once during 1987. More than 92% of all hospitalizations of residents in 1986 were at Saint Joseph's Hospital.[13] The clinic operates the only HMO in the Marshfield Epidemiologic Study Area, and the clinic, HMO, and hospital are all served by a single, salaried medical staff.

From the Marshfield database, we have clinical information on non-HMO as well as HMO patients. To estimate disease occurrence on the basis of medical care utilization, in 1989 we obtained a computerized file of all patients whose current address was then in the Marshfield Epidemiologic Study Area and who had a medical diagnosis of 1 or more of 11 conditions during 1986 through 1988. Both new and old cases of active disease were included.

The file contained several diagnostic records for the same condition for many patients. To calculate annual prevalence of health care utilization for each of the diseases, we eliminated duplicate diagnoses of the same condition.

Population Estimates

To estimate the size of the area population by age and sex for the years 1986 through 1988, we modified zip code–specific population projections for 1989 from a commercial demographics firm.[14] We subtracted age- and sex-specific counts of HMO enrollees in the study area from these figures to derive estimates of the non-HMO population. We calculated disease-specific annual prevalence rates of medical care utilization by payment status, age, and sex. The age and sex distribution of the area population was used to standardize the rates for age and sex by the direct method.[15] We also computed standardized utilization prevalence ratios, analogous to standardized rate ratios, and their 95% confidence intervals.[16]

Results

From 1986 through 1988 the Marshfield Epidemiologic Study Area had an average total population of 47 606, including 26 427 HMO enrollees (56%) and 21 179 non-HMO enrollees (44%). Prevalence of medical care utilization ranged from a low of 3.4 per 10 000 person-years for liver cirrhosis to a high of 201.9 per 10 000 person-years for myocardial infarction (Table 1).

HMO enrollees had significantly higher rates of rheumatoid arthritis, breast cancer, and myocardial infarction than the non-HMO group (Table 2). Conversely, rates of epilepsy and cirrhosis of the liver were significantly lower in the HMO group. In addition, although male HMO enrollees had a significantly higher prevalence of diabetes than nonenrollees, female HMO enrollees had a significantly lower prevalence.

Discussion

Valid disease incidence and prevalence rates are obtainable from a clinical database only if it is population based.[17] This study reports such data from a new epidemiologic laboratory in a rural area served virtually exclusively by a single medical center.

Some of the differences in the patterns of medical care utilization for

specific diseases may be due to selection of HMO coverage by individuals with underlying conditions, such as arthritis, that put them at risk. The differences in the prevalence of medical care utilization for other diseases, however, may reflect different proportions of risk factors, such as alcohol consumption, in the HMO and non-HMO populations.

In addition, some study area residents may travel elsewhere for medical care. If fewer than 10% of the non-HMO enrollees are obtaining their care elsewhere, then our calculations would overestimate the prevalence ratio by less than 9%. There is also potential for error in transcription or extraction from medical records[18,19] and in entering or retrieving information from the computerized database.

The Institute of Medicine recommends that every public health agency regularly and systematically collect, assemble, analyze, and make available information on the health of the community.[20] We encourage physicians and public health officials to incorporate HMO surveillance into their chronic disease control programs on a trial basis. That will permit a balanced assessment of the strengths of HMO data—such as their widespread availability, relative economy, and rich clinical detail—in comparison with their potential weaknesses. □

Acknowledgments

This project was supported by the Bureau of Public Health, Division of Health, Wisconsin Department of Health and Social Services, Madison, under its Sentinel and Chronic Disease Surveillance grant from the Centers for Disease Control and Prevention, Atlanta, Ga.

Selected data from this paper were presented at the 24th Annual Meeting of the Society for Epidemiologic Research, Buffalo, NY, on June 13, 1991, and at the 119th Annual Meeting of the American Public Health Association, Atlanta, Ga, on November 12, 1991.

Thanks are due to Steven K. Broste, MS, Norman A. Desbiens, MD, Larry Hanrahan, MS, William Maurer, MD, Pat Pierce, John Schmelzer, PhD, Cindy Motszko, Brian Weiler, Peggy Peissig, MBA, Laura Brand, Kurt Olsen, MS, Tow Berg, ART, Lisa Lynn, Jill Bulgrin, June Kusta, and Marilyn Daul.

References

1. Last JM. *A Dictionary of Epidemiology.* 2nd ed. New York, NY: Oxford University Press; 1988:125.
2. Centers for Disease Control. Chronic disease reports in the Morbidity and Mortality Weekly Report (MMWR). *MMWR.* 1989; 38(suppl S-1):1–8.
3. Berkelman RL, Buehler JW. Public health surveillance of non-infectious chronic diseases: the potential to detect rapid changes in disease burden. *Int J Epidemiol.* 1990;19: 628–635.
4. Remington PL, Anderson DE, Manering MC, Peterson A, Anderson H. The PRECEDES project: background and methods. *Wis Med J.* 1990;89:695–696.
5. Thacker SB, Berkelman RL. Public health surveillance in the United States. *Epidemiol Rev.* 1988;10:164–190.
6. Greenlick MR, Freeborn DK, Pope CR, eds. *Health Care Research in an HMO: Two Decades of Discovery.* Baltimore, Md: Johns Hopkins University Press; 1988:43.
7. Kelsey JL, Thompson WD, Evans AS. *Methods in Observational Epidemiology.* New York, NY: Oxford University Press; 1986:71–72.
8. Pryor DB, Califf RM, Harrell FE Jr, et al. Clinical data bases: accomplishments and unrealized potential. *Med Care.* 1985;23: 623–647.
9. Kurland LT, Molgaard CA. The patient record in epidemiology. *Sci Am.* 1981;245: 54–63.
10. Chan KA, Felson DT, Yood RA, Walker AM. Incidence of rheumatoid arthritis in central Massachusetts. In: The Society for Epidemiologic Research 25th Annual Meeting Program; June 9–12, 1992; Minneapolis, Minn. Abstract 71.
11. Hansotia P, Broste SK. The effect of epilepsy or diabetes mellitus on the risk of automobile accidents. *N Engl J Med.* 1991;324:22–26.
12. *REZIDE, The National Encyclopedia of Residential ZIP Code Demography.* Alexandria, Va: Claritas; 1985.
13. *Market Share Analyses: Wisconsin* 1986. *Hospitals' Shares of ZIP Code Areas.* Madison, Wis: Wisconsin Division of Health, Center for Health Statistics; 1987: vol 2.
14. Demographic Diskettes [User's Guide]. Fairfax, Va: CACI; 1989.
15. Colton T. *Statistics in Medicine.* Boston, Mass: Little Brown & Co Inc; 1974:47.
16. Checkoway H, Pearce N, Crawford-Brown DJ. *Research Methods in Occupational Epidemiology.* New York, NY: Oxford University Press; 1989:122–130.
17. Dambrosia JM, Ellenberg JH. Statistical considerations for a medical data base. *Biometrics.* 1980;36:323–332.
18. Burnum JF. The misinformation era: the fall of the medical record. *Ann Intern Med.* 1989;110:482–484.
19. Fisher ES, Whaley FS, Krushat WM, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health.* 1992;82: 243–248.
20. Committee for the Study of the Future of Public Health, Division of Health Care Services, Institute of Medicine. *The Future of Public Health.* Washington, DC: National Academy Press; 1988:7.

Popular Films Do Not Reflect Current Tobacco Use

Anna Russo Hazan, PhD, MPH, Helene Levens Lipton, PhD, and Stanton A. Glantz, PhD

ABSTRACT

This study examined trends in tobacco use in a random sample of 2 of the 20 top-grossing US films each year from 1960 through 1990 (62 films). The overall rate of tobacco use did not change. Films continue to portray smokers as successful, attractive White males. Smoking groups became larger, smoking alone declined, hostility and stress reduction were increasingly associated with smoking, and smoking by minor characters increased. Although smoking among elite characters fell, it remained nearly three times as prevalent as in actual population data during the 3 decades. Events involving young people more than doubled. Films do not accurately represent smoking in the United States. (*Am J Public Health*. 1994;84:998–1000)

Introduction

Movies both shape and reflect social values regarding tobacco use. The public believes that there is less smoking than before in movies in the last 5 years.[1] We investigated whether this perception is accurate by studying tobacco use in popular films since 1960.

Methods

Sample Selection

Two feature-length films on videocassette for each year from 1960 through 1990 were randomly selected from the 20 top money-making films each year[2] (62 films). Eight films (13%) were not available on videocassette and were replaced at random with another film from the same year. (Of these eight films, half were from the Disney corporation, which holds back certain films from the video market as a sales strategy.[3])

Data Collection

The title, year of release, rating (G, PG, PG13, R, or X), genre,[4] target audience, and historical era of each of the selected films were recorded. (Ratings for age suitability did not occur until 1968, when the current system was instituted by The Motion Picture Producers and Distributors of America.[5] Films released before 1968 [26%] were unrated. Of the remaining 46 films, 33% [15] were rated G, 52% [24] were rated PG or PG13 [parental guidance suggested], and 15% [7] were rated R [under age 17 not admitted without adult companion]. No X rated [under age 17 not admitted] films appeared in the sample.) The presence of tobacco events in each 5-minute interval of film time was documented upon viewing each videocassette. "Tobacco events" included implied or actual consumption of tobacco, paraphernalia, (ashtrays and matches), talking about tobacco, "no smoking" signs, and tobacco product logos. A per-film reliability ratio of .92 (SD = .07) was computed by comparing the consistency of observations between two coders who viewed the same 15 films independently.

To document tobacco events within each 5-minute interval, 25 variables were coded for every event, including smoker characteristics (e.g., age, sex, role), tobacco characteristics (e.g., product, brand names, tobacco paraphernalia), and scene characteristics (e.g., number of smokers, health messages). Smokers' characteristics were coded when an individual participated in actual smoking or when there was a clear intention to smoke. Intent to smoke was grouped with actual smoking in the analysis, as in prior studies.[6] When only paraphernalia or nonsmoking events were involved, no smoker characteristics were coded.

Statistical Methods

We examined the rate per minute of film at which tobacco events occurred by generating 149 variables from different responses to the original 25 variables (codes and categories of variables are available on request). This rate equaled the total number of times an event occurred divided by the length of the film. Rates were computed because films became shorter by 0.8 minutes per year over the 31-year study period ($P < .01$ by linear regression). We tested for changes over time in the rates of these tobacco events by computing Spearman rank order correlations between the tobacco event rates occurring in each film and the year of release. (The unit of analysis was the film.)

To test for changes decade to decade, we evaluated the characteristics of each film using the original set of 25 variables and cross tabulated the frequency with which each of these characteristics appeared by the decade in which the film was released (1990 was considered

At the time of the study, the authors were with the Institute for Health Policy Studies, Department of Medicine, University of California, San Francisco. Dr Russo Hazan died after the study was completed.

Requests for reprints should be sent to Stanton A. Glantz, PhD, Box 0124, University of California, San Francisco, CA 94143.

This paper was accepted September 23, 1993.

part of the decade of the 1980s); we then analyzed the resulting two-way contingency tables with chi-square tests. (The unit of analysis was the tobacco event.)

Results

The 62 films contained a total of 1505 five-minute intervals (an average of 24.3 intervals per film). Of the 1505 intervals, 36% contained some reference to tobacco; 785 tobacco-related events were recorded for analysis. Tobacco was present and used in 78% (611) of these events and implied in the remaining 22%.

There was virtually no change in the presence of tobacco over time. Of the 149 variables correlated with time, 17 showed significant changes, a number not much more than expected by chance (with $\alpha = .05$).

The motivation of the smoker changed when analyzed by the 3 decades ($P = .03$). Relaxation was the dominant motive (38%, 27%, and 38% in the 1960s, 1970s, and 1980s), respectively. Relaxation motivations included rites of passage (teenage party smoking), ritual (automatic) smoking, and pensive moments.

Thirty-eight percent of tobacco events involved major characters in the 1960s, in comparison with 29% in the 1970s and 26% in the 1980s ($P = .02$). Over the decades, ethnicity changed ($P < .01$). Fewer smoking events involved Whites (81% vs 80% vs 72%), and more involved African Americans (1% vs 4% vs 4%). Gender also changed ($P < .01$). Fewer events involved males (72% vs 79% vs 65%), and more involved females (11% vs 5% vs 15%). Tobacco events involving relatively young people (aged 18 to 29 years) more than doubled (21% vs 22% vs 45%), while the next older group (30 to 45 years) fell by nearly half (39% vs 33% vs 21%) ($P < .01$).

Occasionally (7% of tobacco-related events), smoking was held in contempt, mostly expressed as ridicule. Talk about tobacco was generally positive and inferred pleasure in the act (3% of events over time). These patterns did not change over time.

The size of the group involved in the tobacco events changed, with less smoking in small groups (1 or 2 people) and more in larger groups (16 or more people) ($P < .01$). There was a change in interaction quality ($P < .01$). Congenial interactions declined, and hostile situations increased. There was a small but statistically significant change in the presence of health messages over the decades (2% vs 1% vs 4%) ($P = .05$). Business activities (36%) consistently were the most popular smoking context, while smoking alone declined significantly (13% vs 9% vs 4%) ($P < .01$).

Some smoking behaviors varied over the decades. For example, the smoking level for those 46 to 60 years of age increased from the 1960s (41%) to the 1970s (45%) but decreased in the 1980s (34%) ($P < .01$). This same phenomenon was true for group size "over 16" (15% vs 23% vs 12%) ($P < .01$). Social contexts and motivation showed similar patterns.

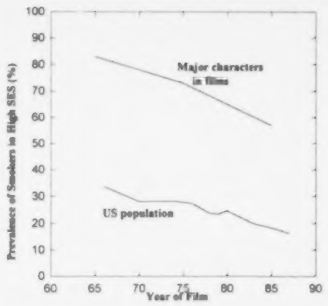
Although the overall rate of smoking was stable over the 3 decades, the prevalence among major elite characters fell over time. Smoking among elites, however, remained nearly three times as prevalent as in the actual population (57% vs 19% in the 1980s) (Figure 1).

Discussion

Films present a smoker who is typically White, male, middle class, successful, and attractive, a movie hero who takes smoking for granted. As in tobacco advertising, smoking in the movies is associated with youthful vigor, good health, good looks, and personal/professional acceptance.[7] There was no significant decline in overall tobacco use over the years 1960 through 1990, but changes in tobacco presentation were found. Smoking became less personal and more social, suggesting that smoking had more populist appeal. There was an increase in the amount of hostility associated with smoking, which may be attributed to the increase in violence in films and the popularity of the antihero. Smoking was used as an anxiolytic, reinforcing its "healthful" properties as a stress reducer, a finding consistent with tobacco industry efforts to promote tobacco use as an alternative intervention to improve mental health.[6]

To some extent, the movies included in this study reflected the reality that better educated, more affluent groups were smoking less over time. While the prevalence of smoking among major characters with high socioeconomic status has decreased, it is still almost three times as prevalent as in comparable population members and has always represented a majority behavior. Another measure of changing social expectations was that ashtrays almost disappeared from movies. The surgeon general has observed that the "ashtray is following the spittoon into oblivion."[8] Films have mirrored this observation. We observed a gradual, consistent decline in the presence of ashtrays (Figure 2).

Movie producers state that they lack established policies about using cigars or cigarettes in their productions other than the overall intention to stay truthful to the characters.[9] They may be using smoking as a means of enhancing the character of a figure, whereas previously, smoking was so prevalent that no such distinction could



Note. The percentage of major characters with high socioeconomic status (equivalent to college graduates based on their occupational status) was calculated and compared with the US population data on smoking by educational status.[8]

FIGURE 1—Smoking prevalence among high socioeconomic status (SES) major characters vs that of comparable members of the US population.

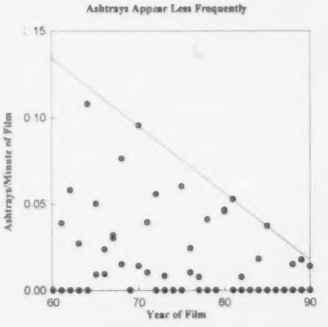


FIGURE 2—Rates at which ashtrays appeared through the period of the study.

be made. Nonetheless, films reinforce misleading images and overstate the normalcy of smoking, which may encourage children and teenagers, the major movie audience, to smoke. □

Acknowledgments

This study was supported by the Cigarette and Tobacco Surtax Fund of the State of California through the Tobacco-Related Disease Research Program of the University of California (award 1RT 520).

Special thanks to Patricia Madden, MPH, of the Prevention Research Center, a project of the Pacific Institute for Research and Evaluation, Berkeley, Calif, for her support and assistance.

References

1. Burns D, Pierce JP. *Tobacco Use in California 1990–1991.* Sacramento, Calif: Department of Health Services; 1992.
2. *Variety: The Show Business Annual.* 1960–1991;54–85.
3. Disney's animated films in short supply by design. *San Francisco Chronicle.* November 2, 1991.
4. *The Video Log.* New York, NY: Trade Services Corp; 1991.
5. Konizsberg I. *Complete Film Dictionary.* New York, NY: New American Library; 1987.
6. Breed W, Defoe JR. Drinking and smoking on television. *J Public Health Policy.* 1984;5: 27–270.
7. *U.S. Federal Trade Commission Staff Report on the Cigarette Advertising Investigation.* Washington, DC: US Federal Trade Commission; 1981.
8. *Reducing the Health Consequences of Smoking: Twenty-five Years of Progress.* Rockville, Md: US Dept of Health and Human Services, Centers for Disease Control; 1989.
9. Seibel DS. Unholy smokes. *San Francisco Chronicle.* June 14, 1992:43.

The Resurgence of Tuberculosis in New York City: A Mixed Hierarchically and Spatially Diffused Epidemic

Deborah Wallace, PhD

ABSTRACT

For New York City from 1978 to 1990, plotting the 3-year running averages of citywide new tuberculosis cases against the middle year yielded an S-shaped curve, with the inflection point at 1983 between early slow and late rapid rise. The inflection in the S curve appears to be associated with hierarchical establishment of secondary epicenters, and the phase of rapid rise in new cases seems to be associated with spatial diffusion from both the primary and secondary epicenters. (*Am J Public Health.* 1994;84:1000–1002)

Introduction

Beginning in the mid-1970s, tuberculosis in New York City edged out from under control as housing overcrowding and residential instability grew in the face of massive low-income housing destruction. This process and aspects of the epidemic that became manifest in 1979 have been described previously.[1]

Medical geographers have developed a taxonomy of contagious spread: hierarchical diffusion, spatial diffusion, and network diffusion.[2,3] Spatial diffusion is the simple geographic spread from an epicenter. Network diffusion refers to a spread via personal and business networks. Hierarchical diffusion is the hardest to understand: the spread from a major urban center where the disease has a very high incidence and prevalence to subordinate communities. The order of spread depends on both the frequency of comings and goings between the primary epicenter and the subordinate communities and the population and population densities of the subordinate communities.[2,3] Once established in these secondary epicenters, the disease spatially diffuses as from the primary epicenter. In this paper, the dynamics of the New York City tuberculosis resurgence are analyzed in light of hierarchical and spatial diffusion.

Materials and Methods

All tuberculosis data were taken from the annual reports of the New York City Tuberculosis Bureau. The bureau's reported new cases and incidences (new cases per 100 000 population) were used. The denominator for calculation of incidence was the last 10-year census population. For 1978 through 1990, the 3-year running averages of the annual new tuberculosis cases were plotted against the middle year of the 3 averaged years.

The following health districts were mapped by year from 1978 through 1990: the districts representing the top quintile for incidence (new cases per 100 000 population) and all other districts with an incidence above the 1978 citywide incidence. The number of cases per square mile was calculated for the top-quintile districts by year from 1978 through 1990 by dividing the number of new cases by the district area in square miles, as provided by the New York City Department of Health.

The author is with Public Service Projects, Consumers Union, Yonkers, NY.

Requests for reprints should be sent to Deborah Wallace, PhD, Public Service Projects, Consumers Union, 101 Truman Ave, Yonkers, NY 10703.

This paper was accepted November 12, 1993.

Results

The graph of the 3-year running average of total citywide new cases vs the middle year of the 3 averaged years shows a classical S shape with the inflection point at 1983 (Figure 1).

The geographic changes over time in incidence are important to analyze to delineate whether this 1982 through 1984 period has characteristics different from the preceding 1978 through 1981 period.

Figure 2 maps the districts of elevated incidence for 1978, 1982, and 1990. By 1982, districts in the top quintile had spread into Brooklyn and the Bronx, whereas in 1978, only one health district outside of Manhattan had ranked in the top quintile (Fort Greene in Brooklyn). These maps also show the spread in elevated incidence in relationship to the top-quintile districts. As the top-quintile districts became well established in both Brooklyn and the Bronx, more districts in these boroughs attained elevated incidence until, by 1990, the entire Bronx was above the 1978 citywide incidence, as were all health districts but one in Brooklyn.

Table 1 displays the number of new cases per square mile (new case density) for the top quintile by incidence rate and year. Although Manhattan routinely had district case densities of more than 25 new cases per square mile during each year of the study period, neither Brooklyn nor the Bronx had a single district with a case density of 25 per square mile or more until 1982. By 1984, both Brooklyn and the Bronx each regularly had at least one health district with 25 or more new cases per square mile. By 1990, five of the six

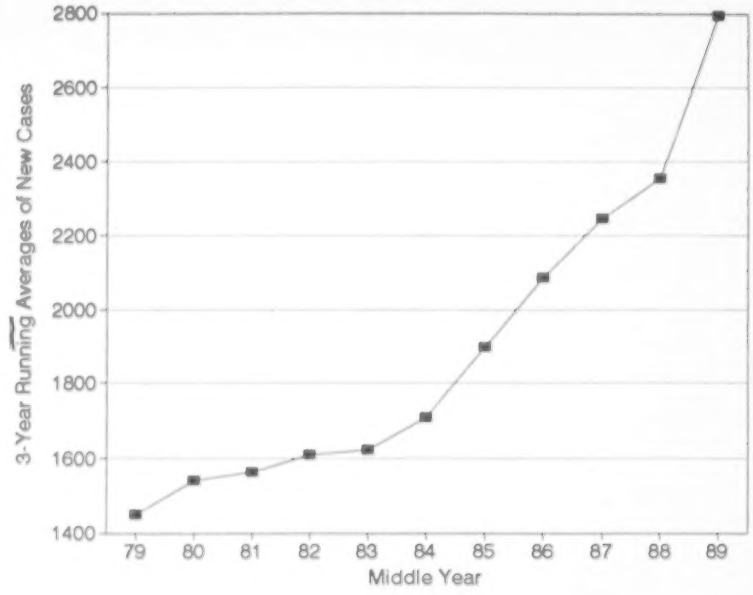
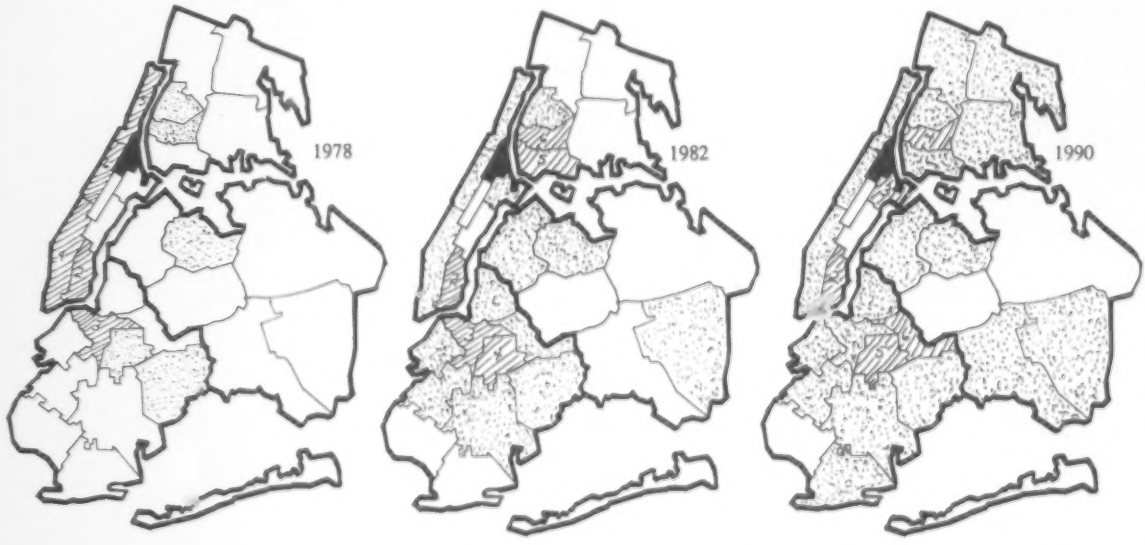


FIGURE 1—Three-year running average of citywide annual new tuberculosis cases vs the middle year of the 3 averaged years.



Note. Black = top district; striping = remaining quintile districts 2–6 as numbered; stippling = "over 1978" district. Staten Island (Borough of Richmond) has been omitted and is not in the top quintile or the "over 1978" group.

FIGURE 2—Maps of the top quintile and of the additional districts with incidences above the 1978 citywide incidence.

TABLE 1—Number of New Tuberculosis Cases per Square Mile: Top-Ranking Districts, by Annual Incidence Rate, New York City, 1978 through 1990

Year	District 1	2	3	4	5	6
1978	**43.10 (CH)**	**29.34 (LE)**	*25.26 (FG)*	**16.22 (LW)**	**28.08 (WH)**	**39.46 (Ri)**
1979	**50.60 (LE)**	**42.06 (CH)**	**21.50 (LW)**	*22.99 (FG)*	**42.99 (RI)**	*21.27 (Be)*
1980	**44.68 (LE)**	**50.37 (CH)**	**19.14 (LW)**	*23.96 (FG)*	*21.02 (Be)*	**36.51 (RI)**
1981	**49.33 (CH)**	**42.53 (LE)**	*24.44 (Be)*	*22.67 (FG)*	**33.00 (WH)**	**32.49 (EH)**
1982	**65.94 (CH)**	**34.45 (LE)**	*21.70 (FG)*	*21.12 (Be)*	12.04 (MH)	25.55 (Mo)
1983	**69.06 (CH)**	**36.61 (LE)**	*21.37 (FG)*	*22.73 (Be)*	12.52 (MH)	25.55 (Mo)
1984	**57.63 (CH)**	**43.87 (LE)**	*25.18 (Be)*	**18.95 (LW)**	*20.73 (FG)*	26.97 (Mo)
1985	**70.10 (CH)**	**45.49 (LE)**	*29.58 (Be)*	*26.88 (FG)*	32.65 (Mo)	*16.35 (Bu)*
1986	**82.04 (CH)**	**55.99 (LE)**	*35.20 (Be)*	*32.71 (FG)*	37.38 (Mo)	**48.06 (EH)**
1987	**81.52 (CH)**	19.84 (MH)	*40.09 (Be)*	**53.30 (LE)**	36.91 (Mo)	**56.16 (WH)**
1988	**98.66 (CH)**	24.09 (MH)	**56.53 (LE)**	42.11 (Mo)	*37.64 (Be)*	**60.93 (EH)**
1989	**105.22 (CH)**	**63.79 (LE)**	*45.46 (Be)*	50.15 (Mo)	*38.86 (FG)*	17.24 (MH)
1990	**142.79 (CH)**	**88.83 (LE)**	**100.87 (EH)**	61.98 (Mo)	*54.75 (Be)*	*29.82 (Bu)*

Note. Manhattan districts appear in boldface, and Brooklyn districts appear in italics. The remaining districts are those of the Bronx. CH = Central Harlem, LE = Lower East Side, FG = Fort Greene, LW = Lower West Side, WH = Washington Heights, Ri = Riverside, Be = Bedford, EH = East Harlem, MH = Mott Haven, Mo = Morrisania, Bu = Bushwick.

top districts had new case densities above 50 per square mile. Top-quintile districts were established in Brooklyn and the Bronx at around the same time as very high spatial concentrations of new cases occurred. Until the Bronx and Brooklyn districts of Morrisania and Bedford reached the density of 25 new cases per mile per year, they did not show annual increases in new case density or incidence.

Discussion

Our benchmark year of 1978, the year of lowest citywide tuberculosis incidence, had remarkable characteristics: the top-ranking health district was not much higher in incidence than the second-ranking district and lay along the same line as the others in the rank vs incidence plot (D. Wallace, unpublished paper); all but one of the top-quintile districts were in Manhattan; and only five districts other than those in the top quintile exceeded the 1978 citywide average. Outside the top quintile, the health districts' new case densities were below 25 per square mile in 1978.

The 1982 through 1984 period formed the inflection of the plot of the 3-year running averages of new cases vs year as the pattern changed from small to large annual increases. The maps of the districts with elevated incidence show that 1982 through 1984 was a period marked by the first establishment of top-quintile districts in both Brooklyn and the Bronx. In addition, in 1982 and 1983, the Morrisania district's new case density rose above 25 per square mile and continued to rise; in 1984, a Brooklyn district, Bedford Stuyvesant, rose above the 25 new cases per square mile density and was joined in 1985 by Fort Greene. Thus, the inflection period features both the spread of the high-ranking incidence districts over three boroughs (instead of being highly concentrated in one) and centers of high case density in three boroughs instead of in one. By 1985, a full system of hierarchical and spatial diffusion had developed and underlay the exponential phase of the epidemic, the postinflection phase of large annual increments. Cliff et al.[2] clearly associated the S-shaped curve of the kind seen for the 3-year running averages with a mixed hierarchical/spatial diffusion. Here, a possible mechanism for the S shape is revealed: first, spatial diffusion from one or two major epicenters (slow early rise); second, hierarchical establishment of secondary epicenters (inflection phase); and third, spatial diffusion from both the primary and secondary epicenters (rapid postinflection rise).

Because a main epicenter and auxiliary epicenters were already established in 1985, control of the disease at the population level will be much more complicated than it would have been before 1982, when only the main epicenter determined the spread and rise of incidence and when housing overcrowding was at a more manageable level. In 1990, new case density in each of these secondary epicenters was above 50 cases per square mile (the case density of the Lower East Side and/or Central Harlem since 1979, when these two districts were the primary epicenters). The data here hint that 25 new cases per square mile is the functional threshold for a secondary epicenter outside of Manhattan and that the establishment of the hierarchical system with primary and secondary epicenters contributed to the change from a slow to a rapid increase in citywide annual new cases. □

References

1. Wallace R, Wallace D. Origins of public health collapse in New York City: the dynamics of "planned" shrinkage, contagious urban decay, and social disintegration. *Bull N Y Acad Med.* 1990;66:391–434.
2. Cliff AD, Haggett P, Ord JK, Versey GR. *Spatial Diffusion: An Historical Geography of Epidemics in an Island Community.* New York, NY: Cambridge University Press; 1981.
3. Abler R, Adams JS, Gould P. *Spatial Organization: The Geographer's View of the World.* Englewood Cliffs, NJ: Prentice Hall; 1971.

Temporal Trends in the Socioeconomic Gradient for Breast Cancer Mortality among US Women

Diane K. Wagener, PhD, and Arthur Schatzkin, MD, DrPh

ABSTRACT

Temporal trends in breast cancer mortality among US women were examined for 1969 through 1989 by age, race, and county-level socioeconomic status (SES). The mortality ratio for high- relative to low-SES counties declined significantly among women 25 to 44, 45 to 64, and more than 65 years of age, respectively, from 1.13 to 0.96, 1.32 to 1.19, and 1.48 to 1.26. The narrowing of mortality occurred among Whites and, to a lesser extent, Blacks. A relative increase in either breast cancer incidence among women in lower SES counties or improved survival among women in higher SES counties (reflecting greater use of screening and treatment) could account for this relative worsening of breast cancer mortality among lower SES women in lower SES counties. (*Am J Public Health.* 1994;84:1003–1006)

Introduction

Breast cancer kills more than 175 000 women throughout the world each year.[1] Among women in the United States, this malignancy accounted for an estimated 46 300 deaths in 1992.[2] Breast cancer mortality, in contrast to most other causes of death,[3] is greater among women having higher, as compared with lower, socioeconomic status (SES).[4,5]

Death statistics in the United States are routinely reported by gender and age but not by SES. Little is known, therefore, about how breast cancer mortality in this country has changed over time for women in different social classes. Because a woman's SES reflects to some degree her risk behaviors and exposure to possible carcinogens,[3–8] as well as her use of screening and treatment services,[3,8–10] investigation of mortality trends by SES may be valuable. In this paper, county-level data are used to examine the relation of SES indicators to breast cancer mortality over the last 2 decades among Black and White women in the United States.

Methods

County socioeconomic data (median family income and percentage of persons 25 years of age or older with at least a high school education) were obtained from the Area Resource File[11] for both the 1970 and 1980 census years. Other studies of breast cancer incidence[4,12–14] and mortality[6] have used these census-based socioeconomic indicators with regions ranging in size from census tract (or block groups) to the county level. Counties (n = 3075) were assigned to quintiles based on the rank of the value for each variable (Table 1). Because less than 15% of the population lived in the two lowest quintile counties, these quintiles were combined to minimize the variance of the ratio estimates.

County-based mortality data, obtained from the Compressed Mortality File,[15–17] were available on Whites separately for 1968 through 1989 and on Blacks separately for 1979 through 1989. Population estimates provided by the National Cancer Institute and death records obtained from the detailed mortality files of the National Center for Health Statistics were used to obtain data for Blacks in 1969 through 1972. Deaths of women 25 years of age or older with an *International Classification of Diseases* (ICD), Eighth Revision (adapted) or Ninth Revision, code of 174 were selected.

Data were combined over several pericensal years. Because the ICD version changed between 1978 and 1979, four (rather than five) pericensal years were combined (i.e., 1969 through 1972 and 1979 through 1982). Also, data from the most recent years (1987 through 1989) were combined.

Data from both files were merged (e.g., 1969 through 1972 deaths and population data were combined with the 1970 SES variables). Because county-level 1990 SES data were not yet available, the 1987 through 1989 mortality data were combined with 1980 SES data.

Statistics include age-specific rates (total number of age-specific deaths over the time period divided by total age-specific person-years); age-adjusted death rates (adjusted to the 1970 US age distribution[18]); standard errors or 95% confidence intervals (CIs) for rates and ratios[18]; and significance of trends of risk ratios.[19]

Results

Breast cancer mortality among US women aged 25 and over rose 3.0% from 1969 through 1972 to 1987 through 1989

Diane K. Wagener is with the Office of Analysis, Epidemiology, and Health Promotion, National Center for Health Statistics, Hyattsville, Md. Arthur Schatzkin is with the Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Md.

Requests for reprints should be sent to Diane K. Wagener, NCHS, Rm 750, 6525 Belcrest Rd, Hyattsville, MD 20782.

This paper was accepted October 1, 1993.

Note. The views expressed are the authors' and do not necessarily represent the policies of the authors' agencies.

TABLE 1—Distribution of Socioeconomic Variables, by Quintiles of Counties

Quintile	Median Family Income[a] 1969, $	1979, $	Total 1979 Population, %	Adult High School Graduates, % 1970	1980	Of Total 1980 Population
1	5 788	13 750	4.9	31.7	46.8	5.5
2	5788–6859	13 750–15 629	7.7	31.7–41.8	46.8–57.1	11.1
3	6860–7875	15 630–17 320	11.9	41.9–49.5	57.2–64.2	17.0
4	7876–8904	17 321–19 526	18.9	49.6–55.9	64.3–69.9	29.8
5	8 904	19 526	56.7	55.9	69.9	36.6
Minimum value	2 407	7 170		11.6	25.1	
Maximum value	18 333	33 711		88.2	95.3	

[a]The poverty level for a family of four was $3743 in 1969 and $7412 in 1979.

TABLE 2—Breast Cancer Mortality Rates among US Women, by Age and Race

Race/Age, y	Rate (95% Confidence Interval) 1969–1972	1979–1982	1987–1989
White			
All adults	49.6 (49.3, 49.9)	49.4 (49.1, 49.7)	50.7 (50.4, 51.0)
25–44	11.1 (10.9, 11.3)	9.8 (9.6, 10.0)	9.5 (9.3, 9.7)
45–64	64.4 (63.8, 64.9)	62.2 (61.7, 62.7)	61.1 (60.5, 61.7)
65+	110.9 (109.9, 111.9)	117.1 (116.2, 118.0)	127.2 (126.2, 128.2)
Black			
All adults	43.7 (42.8, 44.6)	49.5 (48.7, 50.3)	56.7 (55.7, 57.7)
25–44	13.1 (12 5, 13.7)	13.9 (13.3, 14.5)	15.9 (15.2, 16.6)
45–64	53.5 (51.9, 55.1)	65.8 (64.2, 67.4)	74.0 (72.1, 75.9)
65+	76.8 (73.9, 79.7)	100.8 (98.0, 103.6)	118.5 (115.3, 121.7)
All			
All adults	49.2 (48.9, 49.5)	49.0 (48.7, 49.3)	50.7 (50.4, 51.0)
25–44	11.5 (11.3, 11.7)	10.2 (10.0, 10.4)	10.1 (9.9, 10.3)
45–64	63.8 (63.3, 64.3)	62.0 (61.5, 62.5)	61.6 (61.0, 62.2)
65+	108.8 (107.9, 109.7)	114.9 (114.1, 115.7)	125.2 (124.5, 126.1)

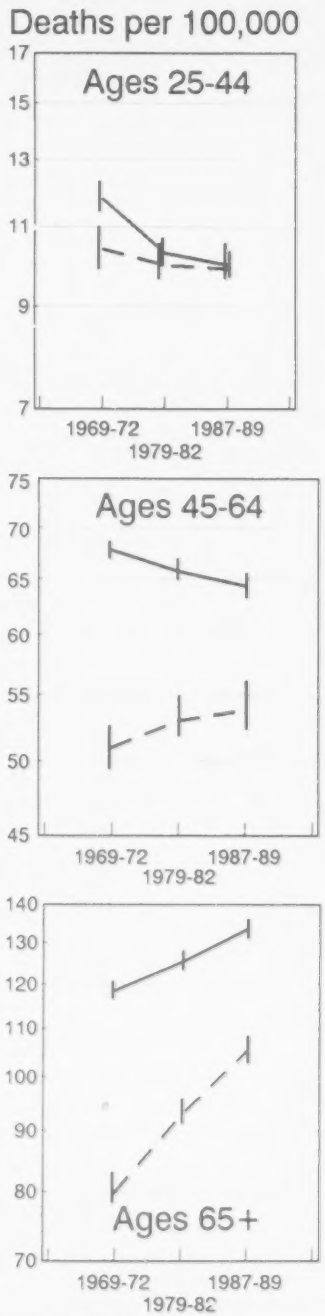
Note. Mortality rates are per 100 000 women. Mortality was age-adjusted to the 1970 US population.

(Table 2). Different trends are evident if the death statistics are broken down by age and race. Among Black women, mortality climbed 21.4% for those 25 to 44 years of age, 38.3% for those 45 to 64, and 54.3% for those over 64. Among Whites, mortality declined 14.4% among younger women and 5.1% among middle-aged women but increased 14.7% among older women.

The relative worsening of mortality for Blacks is reflected in trends for the Black–White mortality ratio. In the interval from 1969 through 1972 to 1987 through 1989, this ratio increased from 1.18 (95% CI = 1.12, 1.24) to 1.67 (95% CI = 1.59, 1.75) for women aged 25 to 44 years, from 0.83 (95% CI = 0.80, 0.86) to 1.21 (95% CI = 1.18, 1.24) for women 45 to 64, and from 0.69 (95% CI = 0.66, 0.72) to 0.93 (95% CI = 0.90, 0.95) for women 65 or older.

Figure 1 depicts breast cancer mortality trends among quintiles defined by county median family income for all races combined (see also Table 3). Despite the opposite trends of mortality in younger and older women, the SES differentials narrowed for both age groups. The SES differential also narrowed among middle-aged women. Education quintiles provide similar but slightly attenuated results both within time periods and over time (data are available from the authors on request).

Table 3 also presents, race, age-, income-, and time period–specific mortality as well as ratios of mortality among the



Lowest Quintiles (1 and 2) - - - -
Highest Quintile (5) ——
Note. Mortality rates (and confidence intervals, shown as vertical bars) are shown for 1969–1972, 1979–1982, and 1987–1989 among women 25–44 years old, 45–64 years old, and 65 years old or older. Death rates were age-adjusted to the 1970 US population.

FIGURE 1—Breast cancer mortality, by quintile of median family income (multiyear averages).

highest quintile populations and the two lowest quintile populations. Among Whites, all of the changes in ratios were significant. For Black women, the high–low mortality ratio was unchanged among the younger women and decreased somewhat for the other age groups, in part because of a substantial breast cancer mortality increase in the poorer counties.

Discussion

Our study shows a narrowing of the direct SES differential in the breast cancer gap over the past 2 decades among US women for each age and race group (except the youngest Black women). Breast cancer mortality, which has been rising among women 65 years of age and older, has increased more rapidly among those in lower SES counties. Among women 45 to 64, breast cancer mortality has been declining in those women in higher SES counties but rising in their lower SES counterparts. Finally, women 25 to 44 have experienced a fall in breast cancer death rates over the last 2 decades, with the decline being more rapid in those in higher SES counties. Earlier data (through 1971) from the United Kingdom suggest a similar shrinking of the social class gradient in breast cancer mortality.[7]

These analyses were areal (or "ecologic"). County-level SES may be a proxy for *individual-level* measures (e.g., income, education, and type of health insurance) or a reflection of *county-level* characteristics (e.g., population density, pollution, and availability of screening and treatment services). Indeed, there may be a mix of individual- and county-level factors at work. Further studies using census tracts or blocks or other areas smaller than the county for ascribing SES might reduce the classification error known to exist in ecologic analyses.[20–22]

Data from other studies suggest that these observed SES trends in breast cancer mortality can be explained by changes in both incidence and survival. Cross-sectional ecologic analyses by SES indicate that breast cancer incidence is greater among women of higher SES.[4,23] However, between 1974 and 1984, the incidence in the Seattle region increased more rapidly among women living in low-income census tracts.[14] Use of both mammography[9] and breast examination[9,10] has been shown to be lower and to have increased more slowly between 1974 and 1984 among poorer or less educated women in the United States. Cross-sectional data indicate that lower SES

TABLE 3—Breast Cancer Mortality Rates among US Women, by Age, Race, and Median Family Income Quintiles

Age, y	Race	Income Quintile	Rate (95% Confidence Interval) 1969–1972	1979–1982	1987–1989
25–44	White	Q5 (high)	11.6 (11.4, 11.8)	10.1 (9.9, 10.3)	9.5 (9.3, 9.7)
		Q4	10.4 (9.8, 11.0)	9.3 (8.9, 9.7)	9.3 (8.9, 9.7)
		Q3	10.3 (9.5, 11.1)	9.7 (9.1, 10.3)	9.9 (9.3, 10.5)
		Q2	9.7 (8.7, 10.7)	9.3 (8.5, 10.1)	9.2 (8.4, 10.0)
		Q1 (low)	9.7 (8.7, 10.7)	9.4 (8.4, 10.4)	9.8 (8.8, 10.8)
		Q5:Q1, 2	1.20 (1.09, 1.31)	1.07 (1.00, 1.14)	1.01 (0.94, 1.08)
		Trend: $P = .03$			
	Black	Q5 (high)	12.5 (11.7, 13.3)	14.2 (13.2, 15.2)	15.7 (14.7, 16.7)
		Q4	14.0 (12.2, 15.8)	13.5 (11.9, 15.1)	16.6 (14.8, 18.4)
		Q3	17.7 (14.8, 20.6)	13.5 (11.7, 15.3)	15.8 (13.6, 18.0)
		Q2	10.8 (8.4, 13.2)	14.0 (12.0, 16.0)	15.8 (13.4, 18.2)
		Q1 (low)	13.5 (11.0, 16.0)	12.4 (10.0, 14.8)	15.1 (12.4, 17.8)
		Q5:Q1, 2	1.03 (0.86, 1.19)	1.06 (0.91, 1.21)	1.01 (0.89, 1.13)
		Trend: $P = .50$			
	All	Q5 (high)	11.8 (11.6, 12.0)	10.4 (10.2, 10.6)	10.0 (9.8, 10.2)
		Q4	11.0 (10.4, 11.6)	9.7 (9.3, 10.1)	10.0 (9.4, 10.6)
		Q3	11.3 (10.5, 12.1)	10.3 (9.7, 10.9)	10.6 (9.7, 11.2)
		Q2	10.3 (9.5, 11.1)	10.1 (9.7, 10.9)	10.2 (9.4, 11.0)
		Q1 (low)	10.5 (9.5, 11.5)	9.9 (9.1, 10.7)	10.8 (9.8, 11.8)
		Q5:Q1, 2	1.13 (1.06, 1.20)	1.04 (0.96, 1.12)	0.96 (0.91, 1.02)
		Trend: $P = .02$			
45–64	White	Q5 (high)	68.2 (67.4, 69.0)	65.6 (64.8, 66.4)	64.7 (63.9, 65.5)
		Q4	62.4 (61.0, 63.8)	58.3 (57.1, 59.5)	58.5 (57.1, 59.9)
		Q3	54.6 (52.8, 56.4)	57.0 (55.4, 58.6)	57.6 (55.8, 59.4)
		Q2	52.4 (50.2, 54.6)	55.3 (53.5, 57.1)	54.2 (52.0, 56.4)
		Q1 (low)	49.9 (47.5, 52.3)	50.4 (48.2, 52.6)	50.6 (48.1, 53.1)
		Q5:Q1, 2	1.32 (1.28, 1.36)	1.23 (1.20, 1.26)	1.23 (1.19, 1.27)
		Trend: $P = .02$			
	Black	Q5 (high)	56.9 (54.7, 59.1)	68.4 (65.9, 70.9)	75.9 (73.2, 78.6)
		Q4	53.0 (49.1, 56.9)	68.2 (64.3, 72.1)	78.5 (73.8, 83.2)
		Q3	49.4 (44.1, 54.7)	68.0 (63.9, 72.1)	74.1 (69.0, 79.2)
		Q2	46.8 (41.7, 51.9)	58.9 (54.4, 63.4)	65.0 (59.7, 70.3)
		Q1 (low)	44.3 (39.6, 49.0)	50.4 (45.3, 55.5)	63.9 (57.2, 70.6)
		Q5:Q1, 2	1.25 (1.14, 1.36)	1.23 (1.14, 1.32)	1.17 (1.09, 1.25)
		Trend: $P = .14$			
	All	Q5 (high)	67.7 (67.1, 68.3)	65.7 (65.0, 66.4)	64.4 (63.6, 65.2)
		Q4	62.1 (60.9, 63.3)	59.2 (58.0, 60.4)	59.9 (58.5, 61.3)
		Q3	54.7 (52.9, 56.5)	58.4 (57.0, 59.8)	59.3 (57.7, 60.9)
		Q2	52.5 (50.5, 54.5)	55.6 (53.8, 57.4)	55.3 (53.3, 57.3)
		Q1 (low)	49.5 (47.3, 51.7)	50.3 (48.3, 52.0)	52.2 (50.8, 54.6)
		Q5:Q1, 2	1.32 (1.28, 1.36)	1.23 (1.20, 1.26)	1.19 (1.15, 1.23)
		Trend: $P = .01$			
65+	White	Q5 (high)	120.3 (118.9, 121.7)	125.3 (124.1, 126.5)	136.2 (134.8, 137.6)
		Q4	105.0 (102.6, 107.4)	109.0 (107.0, 111.0)	121.5 (119.3, 123.7)
		Q3	96.1 (93.2, 99.0)	109.8 (107.4, 112.2)	119.7 (117.0, 122.4)
		Q2	86.0 (82.3, 89.7)	100.9 (98.0, 103.8)	110.2 (106.9, 113.5)
		Q1 (low)	75.6 (71.9, 79.3)	89.9 (86.2, 93.6)	102.6 (98.7, 106.5)
		Q5:Q1, 2	1.48 (1.43, 1.53)	1.30 (1.27, 1.33)	1.27 (1.24, 1.30)
		Trend: $P = .005$			
	Black	Q5 (high)	87.4 (82.9, 91.9)	111.5 (106.8, 116.2)	122.3 (117.2, 127.4)
		Q4	71.5 (65.0, 78.0)	103.5 (97.2, 109.8)	125.6 (118.2, 133.0)
		Q3	69.3 (60.5, 78.1)	98.3 (91.8, 104.8)	121.3 (113.5, 129.1)
		Q2	64.1 (56.1, 72.1)	95.1 (87.7, 102.6)	107.7 (99.3, 116.5)
		Q1 (low)	65.4 (58.1, 72.7)	66.2 (59.3, 73.1)	95.7 (86.3, 105.1)
		Q5:Q1, 2	1.34 (1.21, 1.47)	1.36 (1.26, 1.46)	1.19 (1.10, 1.28)
		Trend: $P = .11$			
	All	Q5 (high)	118.6 (117.4, 119.8)	125.1 (123.9, 126.3)	133.3 (131.9, 134.7)
		Q4	102.8 (100.6, 105.0)	108.1 (106.3, 109.9)	121.1 (118.9, 123.3)
		Q3	94.8 (92.1, 97.5)	108.4 (106.2, 110.6)	119.2 (116.7, 121.7)
		Q2	84.4 (81.3, 87.5)	99.6 (96.9, 102.3)	109.1 (106.2, 112.0)
		Q1 (low)	75.4 (71.7, 79.1)	85.7 (82.8, 88.6)	101.3 (97.8, 104.8)
		Q5:Q1, 2	1.48 (1.43, 1.53)	1.33 (1.30, 1.36)	1.26 (1.23, 1.29)
		Trend: $P = .005$			

Note. Mortality rates are per 100 000 women. Mortality was age-adjusted to the 1970 US population. Q5:Q1, 2 = rate ratio of mortality in highest income quintile (Q5) to population-based weighted average mortality of two lowest income quintiles (Q2 and Q1). Because of the nature of ratio estimates, the ratio for the total population is not a simple average of the ratios for each subpopulation. Trend = test for trend over time of ratios of Q5 to Q1 and Q2 (combined).

women present at a more advanced stage of disease.[24] Whether treatment differentials by SES play a role is unclear. Thus, the diminishing mortality gap over time among women living in high- as compared with low-SES counties may reflect (1) a more rapidly rising incidence of breast cancer among younger women in low-income counties or (2) greater survival among women in high- income, or poorer survival among women in low-income, counties.

Women in lower SES counties appear to be reaching "equity" in breast cancer mortality with those living in higher SES counties. This, unfortunately, is not the kind of SES equity one might hope for. Further investigations of temporal trends in the SES distribution of risk factors (for incidence and survival) might provide additional etiologic leads and direct attention to deficiencies in breast cancer medical services. This would benefit women from all social strata. □

Acknowledgments

The authors wish to thank Drs Larry Kessler and Robert Tarone for their critical comments on an earlier draft of this manuscript.

References

1. *World Health Statistics Annual 1990.* Geneva, Switzerland: World Health Organization; 1991.
2. *Cancer Facts and Figures, 1992.* Atlanta, Ga: American Cancer Society; 1992.
3. Pappas G, Queen S, Hadden W, Fisher G. The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *N Engl J Med.* 1993; 329:103–109.
4. Devesa SS, Diamond EL. Association of breast cancer and cervical cancer incidences with income and education among whites and blacks. *J Natl Cancer Inst.* 1980;65:515– 528.
5. Hakama M, Hakulinen T, Pukkala E, et al. Risk indicators of breast and cervical cancer on ecologic and individual levels. *Am J Epidemiol.* 1982;116:990–1000.
6. Kelsey JL. A review of the epidemiology of human breast cancer. *Epidemiol Rev.* 1979; 1:74–109.
7. Logan WP. *Cancer Mortality by Occupation and Social Class 1851–1971.* London, England: Her Majesty's Stationery Office; 1982. Studies on Medical and Population Subjects 44.
8. Angell M. Privilege and health—what is the connection? *N Engl J Med.* 1993;329: 126– 127.
9. Dawson DA, Thompson GB. Breast cancer risk factors and screening: United States, 1987. *Vital Health Stat [10].* 1990; 172. DHHS publication PHS 90-1500.
10. Makuc DM, Freid VM, Kleinman JC. National trends in the use of preventive health care by women. *Am J Public Health.* 1989; 79:21–26.
11. Stambler HV. The Area Resource File—a brief look. *Public Health Rep.* 1988;103: 184–188.
12. McWhorter WP, Schatzkin AG, Horm JW, Brown CC. Contribution of socioeconomic status to black/white differences in cancer incidence. *Cancer.* 1989;63:982–987.
13. Baquet CR, Horm JW, Gibbs T, Greenwald P. Socioeconomic factors and cancer incidence among blacks and whites. *J Natl Cancer Inst.* 1991;83:551–556.
14. White E, Daling JR, Norsted TL, Chu J. Rising incidence of breast cancer among young women in Washington State. *J Natl Cancer Inst.* 1987;79:239–243.
15. Ingram DD, Gillum RF. The Mortality Surveillance System and its application tothe study of ischemic heart disease mortality. *J Am Stat Assoc.* 1987;82:324–327.
16. Ingram DD, Gillum RF. Regional and urbanization differentials in coronary heart disease mortality in the United States, 1968–85. *J Clin Epidemiol.* 1989;42:857–868.
17. *Public Use Data Tape Documentation: Compressed Mortality File, 1968–1985.* Hyattsville, Md: National Center for Health Statistics; 1989.
18. Kleinman JC. Mortality. In: *Statistical Notes for Health Planners No. 3.* Washington, DC: US Dept of Health, Education and Welfare; 1977. DHEW publication HRA 77-1237.
19. Rothman KJ. *Modern Epidemiology.* Boston, Mass: Little, Brown; 1986:343–344.
20. Alker JR Jr. A typology of ecologic fallacies. In: Doggan M, Rokkan S, eds. *Social Ecology.* Cambridge, Mass: MIT Press; 1969:69–86.
21. Haan M, Kaplan G, Camacho T. Poverty and health: prospective evidence from the Alameda County study. *Am J Epidemiol.* 1987;125:989–998.
22. Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *Am J Public Health.* 1992; 82:703–710.
23. Krieger N. Social class and the black/white crossover in the age-specific incidence of breast cancer: a study linking census-derived data to population-based registry records. *Am J Epidemiol.* 1990; 131:804–814.
24. Wells BL, Horm JW. Stage at diagnosis in breast cancer: race and socioeconomic factors. *Am J Public Health.* 1992;82:1383–1385.

Candidates Sought for Reproductive Health Initiative Awards

The American Medical Women's Association (AMWA) announces its first Reproductive Health Initiative Awards to recognize physicians who have made outstanding contributions to reproductive health for women. The $10 000 national award will be presented to a male or female physician who is a prominent and well-known current reproductive health care provider and who, in the view of AMWA, has made significant contributions to advancing women's reproductive health and has attempted to bring these issues into mainstream medicine.

AMWA will consider candidates who have dedicated personal time and their professional career to providing quality reproductive health care services, including abortion services. Potential awardees will be those who have provided leadership and education and who have promoted reproductive health issues on a national level.

The new award is part of AMWA's Reproductive Health Initiative, a 3-year project to develop and promote curricula that medical schools can use to train medical students in reproductive health. Interested medical schools can participate in the initiative training by holding 1-day seminars. In addition, AMWA will arrange for fourth-year medical students to be trained in reproductive health.

To be considered for the national award, those nominating potential candidates should provide a nomination statement between 50 and 250 words and send it with the nominee's curriculum vitae, nomination form, and other relevant supporting information to AMWA by *Friday, July 22, 1994.* For a national application, contact Susan Eisendrath, MPH, Project Coordinator–RHI, c/o AMWA, 801 N Fairfax St, Alexandria, VA 22314; tel (703) 838-0500.

Five local awards of $1000 each will be granted through AMWA branches. To be considered for a local award, contact AMWA for the name of the president of the branch in your area.

Racial Differences in Cervical Cancer Mortality in Chicago

Elizabeth J. Samelson, MPH, Marjorie A. Speers, PhD, Richard Ferguson, MS, and Charles Bennett, PhD

ABSTRACT

Racial differences in cervical cancer mortality in Chicago were examined. Age-adjusted mortality in Blacks (10.0/100 000) was over twice the rate found in Whites (4.6/100 000). Age-specific rates also showed significant excess mortality among Blacks. After stratification by a group-level defined poverty indicator, the race differential in age-adjusted rates remained significant. The race differential in age-specific rates diminished in the group with more than 30% living below the national poverty level, in contrast to the group with 30% or fewer living below the national poverty level, in whom race differences were more marked. Methodological issues concerning hysterectomy prevalence, Hispanic ethnicity, and social class must be considered with respect to interpretation of these findings. (*Am J Public Health.* 1994;84:1007–1009)

Introduction

Cancer of the cervix is one of the cancer sites most amenable to early detection and treatment, yet it is the second leading cause of cancer mortality for women worldwide[1] and ranks third among cancer deaths nationally for women aged 25 through 44 years.[2] Throughout the last decade, marked excess cervical cancer mortality in Blacks has persisted despite declining rates in both Black and White women.[3]

Chicago's mortality rates for cervical cancer deaths were 14% higher than national rates for Whites and 24% higher for non-Whites from 1968 through 1982.[4] The present study was conducted as part of a federally funded state health department program established to develop cervical cancer prevention strategies. The program was divided into three phases: descriptive epidemiologic research, analytic study, and finally, the design and implementation of a demonstration project. The present study was undertaken to provide a descriptive account of racial differences in cervical cancer mortality among women in Chicago.

Materials and Methods

The study area included 37 central Chicago communities (representing approximately 53% of Chicago's total population) selected from the 1980 census. The far northern and southern portions of Chicago were excluded to reduce the number of patients more likely to utilize suburban or out-of-state health facilities and to include a greater proportion of low-income and minority women.

Estimates of the female population residing in the study area were obtained from the 1980 census, the only available data specific for age, race, gender, and community area. Race was classified as White (including Hispanic) and non-White. Because 93% of the non-White group was Black, this term will be used throughout this report.

Deaths from cervical cancer (*International Classification of Diseases*, Ninth Revision, codes 180.0–180.9) were selected from the state of Illinois death tapes for the period from 1975 through 1984. Average annual mortality rates are presented. Information on poverty status was not available on an individual level for this study. Nevertheless, an effort was made to account for potential confounding by categorizing subjects, at the group level, according to the percentage of residents ($>30\%$, $\leq 30\%$) living below the national poverty level in the community area where each resided. Thirty percent was chosen as the value to dichotomize this poverty variable in order to minimize the number of cells with zero cases and to ensure a reasonable estimate of poverty in the community areas.

Age-standardized rates were calculated by the direct method[5] with the combined population of the 37 community areas as the standard. The z test (two tailed) based on the standard normal distribution was used for statistical testing of race differences in the age-specific rates[6] and the overall age-adjusted rates.[5] (Significance was defined at $P < .05$.) Rate ratios (and 95% confidence intervals [CIs]) comparing mortality in Blacks with that in Whites were calculated with DEPID.[7]

Results

From 1975 through 1984, 600 cervical cancer deaths were reported for the study

At the time of the study, Elizabeth J. Samelson was with the Cervical Cancer Demonstration and Control Project, Illinois Department of Public Health, Chicago, Ill; she is currently a doctoral student in epidemiology at Columbia University, New York, NY. Marjorie A. Speers is with the Division of Chronic Disease Control and Community Intervention, Centers for Disease Control and Prevention, Atlanta, Ga. At the time of the study Richard Ferguson was with the Illinois Department of Public Health; he is currently with the Cook County Bureau of Health Services, Chicago, Ill. Charles Bennett was with the Southern Illinois School of Medicine, Springfield, Ill; he is currently with the Vermont Department of Health, Burlington, Vt.

Requests for reprints should be sent to Elizabeth J. Samelson, MPH, Division of Epidemiology, Columbia University, School of Public Health, 600 W 168th St, New York, NY 10032.

This paper was accepted November 4, 1993.

TABLE 1—Average Annual Mortality Rates,[a] Rate Ratios,[b] and Confidence Intervals (CIs): Cervical Cancer Project, Chicago Study Area,[c] 1975 through 1984

	White		Black				>30% Poverty		≤30% Poverty	
	Rate	No.	Rate	No.	Rate Ratio	95% CI	Rate Ratio	95% CI	Rate Ratio	95% CI
Age group, y										
15–34	0.7	9	1.6	26	2.3	1.1, 4.9	. . .	. . .	2.3	1.0, 5.7
35–54	9.6	68	20.7	184	2.2	1.6, 2.8	1.8	1.0, 3.3	1.7	1.5, 2.4
55–74	12.7	93	21.7	126	1.7	1.3, 2.2	1.2	0.7, 2.1	1.8	1.2, 2.5
75+	13.0	28	46.0	66	3.6	2.3, 5.5	2.2	0.9, 5.6	3.5	1.9, 6.4
Total sample										
Crude rate	5.2	198	8.8	402	1.9	1.6, 2.1	1.6	1.3, 1.9	1.5	1.2, 1.9
Adjusted rate[d]	4.6		10.0		2.1	1.8, 2.5	1.7	1.2, 2.4	1.9	1.5, 2.4

[a]Rates per 100 000. Rates based on fewer than 5 cases were not calculated.
[b]White is reference group.
[c]Study area includes 37 geographically contiguous communities located in central Chicago selected from the 1980 census.
[d]Rates were age adjusted to the total study population by the direct method.

area (Table 1). Almost half the deaths (287) occurred among women younger than 55 years of age. The age-adjusted mortality rate of 10.0 for Black women was more than twice the overall rate of 4.6 for Whites ($P < .001$). A comparison of age-specific rates yielded similar results: Blacks had statistically significant greater mortality at every age level. Mortality rates for Black women were more than 2 times those for White women in the 15–34 and 35–54 age groups, 1.7 times those for White women in the 55–74 age group, and more than 3.5 times those for White women in the 75+ age group. Mortality rates increased significantly with age for both races, but the greatest increases occurred among Black women.

The race differential in the age-adjusted mortality rates remained statistically significant after stratifying by poverty group. The Black–White rate ratio was 1.7 (CI = 1.2, 2.4) for the >30% poverty group and 1.9 (CI = 1.5, 2.4) for the ≤30% poverty group (Table 1).

Within the community areas of ≤30% poverty, mortality was significantly greater in Blacks in all but the youngest age group (rates not shown). Although the rates for Blacks were greater than the rates for Whites at every age level in the >30% poverty group (rates not shown), none were statistically significant.

Discussion

The marked excess of cervical cancer deaths among Black women in Chicago described in this report is consistent with national data demonstrating more than a twofold increase in mortality rates for Black women.[3] Several methodological issues must be considered with respect to the interpretation of our findings.

The rates presented here included women with prior hysterectomies; consequently, mortality was underestimated in this study. The effect on the rate ratios is prohibitively complex to predict because of potential age, race, and socioeconomic differences in hysterectomy prevalence. Results from the Behavioral Risk Factor Surveillance System, however, did show that overall frequency of prior hysterectomy was the same for Whites, Blacks, and Hispanics in a national sample of women; lower income and less education, however, were related to increased prevalence of hysterectomy.[8]

A second limitation of this study concerns the validity of using census denominators to enumerate the Whites and non-Whites in the study population. Because information on ethnicity was not included in mortality data at the time of this analysis, rates for Hispanics could not be calculated separately from those for Whites. As a result, the rates for Whites in this study may have been overestimated (and the rate ratio underestimated) because Hispanic women have been identified as a high-risk group for cervical cancer.[9–11] Alternatively, if the number of non-Whites was underestimated, the comparisons between Whites and non-Whites may have exaggerated the race differential in mortality.

The >30% poverty group comprised a greater proportion of Hispanics to Whites than did the ≤30% poverty group.[12] The inclusion of Hispanic women in the White category may, in part, account for the insignificant race difference observed in the group with >30% poverty in contrast to the ≤30% poverty group, in whom race differences were significantly more marked. Two other explanations for this finding can also be given. First, the stresses associated with poverty may present barriers to obtaining preventive health services for both Blacks and Whites residing in the most economically disadvantaged communities.[13] Alternatively, whereas social class may be more similar between Whites and Blacks in the >30% poverty group, social class may be more disparate between Whites and Blacks in the ≤30% poverty group. Although our data can neither confirm nor reject these explanations, other studies have shown that race differences in Pap smear screening[14] and incidence of invasive cervical cancer[16] are more marked in higher than in lower socioeconomic areas.

Because poverty status was defined at the group level (according to the percentage of residents living below the national poverty level in the community where each study subject resided), no definitive conclusions may be applied on an individual basis. However, because community areas in Chicago tend to be particularly homogeneous with respect to socioeconomic status, significant misclassification was unlikely. Moreover, group-level definitions of economic status have

been used in other studies of race differences in cervical cancer, which reported results similar to those in studies that used individual indicators.[15,16]

Investigators have suggested for many decades that disparities in screening, in part, account for the excess cervical cancer mortality in Black women.[14,17–20] More recent data from the Behavioral Risk Factor Surveillance System, however, showed that a higher proportion of Blacks than Whites obtained Pap smears.[21] Results from the 1987 National Health Interview Survey demonstrated that poor White women were more likely to have never had a Pap smear than poor Black women.[22] These findings suggest that patient treatment and follow-up also play an important role in the excess mortality in cervical cancer seen in Blacks. Improved recording of Pap smears, specifically through efficient registries, has been recommended to provide better tracking of women with abnormal cytology reports and to promote provider adherence to screening and treatment guidelines.[23,24]

In addition to disparities in screening practices, the race differential in cervical cancer mortality can also be viewed as a function of the less favorable survival experience seen among Blacks. Black women are not only less likely to be diagnosed with localized cervical lesions, but are less likely than Whites to survive at the same stage of diagnosis.[18,25]

In summary, our findings are consistent with other studies reporting a nearly twofold race differential in cervical cancer mortality between White and Black women. The results from this study support intervention efforts focused on reducing cervical cancer mortality among Black women in Chicago. Future studies might attempt to improve measurement of social class to better assess its role in cervical cancer mortality. □

Acknowledgments

Funding was supplied through Centers for Disease Control and Prevention cooperative agreement U50-CCU50221.

References

1. Munoz N, Bosch FX. Epidemiology of cervical cancer. In: Munoz N, Bosch FX, Jensen DM, eds. *Human Papillomavirus and Cervical Cancer.* Oxford, England: International Agency for Research on Cancer; 1989. IARC scientific publication 94.
2. Devesa SS. Cancer mortality, incidence and patient survival among American women. *Women and Health.* 1986;11:7–22.
3. Centers for Disease Control. Black-white differences in cervical cancer mortality—United States, 1980–1987. *MMWR.* 1990;39: 245–248.
4. Mallin K, Haenszel W. *A Review of Cancer Mortality in Chicago and Chicago Community Areas 1968–1982.* Chicago, Ill: Illinois Cancer Council; 1986:18.
5. Kahn HA, Sempos CT. *Statistical Methods in Epidemiology.* New York, NY: Oxford University Press; 1989:87–92.
6. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley & Sons Inc; 1981:23.
7. Sullivan KM, Foster DA. *DEPID, A Program for Stratified and Standardized Analysis, Version 2.1.* 1987.
8. Kjerulff K, Langenberg P, Guzinski G. The socioeconomic correlates of hysterectomies in the United States. *Am J Public Health.* 1993;83:106–108.
9. Baquet CR, Ringen K, Pollack ES, et al. *Cancer among Blacks and Other Minorities: Statistical Profiles.* Bethesda, Md: National Cancer Institute; 1986. NIH publication 86-2785.
10. Peters RK, Mack TM, Bernstein L. Parallels in the epidemiology of selected anogenital carcinomas. *J Natl Cancer Inst.* 1984;72: 609–615.
11. Peters RK, Thomas D, Hagan DG, Mack TM, Henderson BE. Risk factors for invasive cervical carcinoma among Latinas and non-Latinas in Los Angeles County. *J Natl Cancer Inst.* 1986;77:1063–1077.
12. *Chicago Statistical Abstract.* Chicago, Ill: City of Chicago, Department of Planning; 1980.
13. Navarro V. Race or class or race and class: growing mortality differentials in the US. *Int J Health Serv.* 1991;21:229–235.
14. Kleinman JC, Kopstein A. Who is being screened for cervical cancer? *Am J Public Health* 1981;71:73–76.
15. Devesa SS, Diamond EL. Association of breast and cervical cancer incidences with income and education among whites and blacks. *J Natl Cancer Inst.* 1980;65:515–528.
16. Mandelblatt J, Andrews H, Kerner J, Zauber A, Burnett W. Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type. *Am J Public Health.* 1991;81:646–649.
17. Rochat RW. The prevalence of cervical cancer screening in the United States in 1970. *Am J Obstet Gynecol.* 1976;125:478–485.
18. Breslow LB, Hochstim JR. Sociocultural aspects of cervical cytology in Alameda County, California. *Public Health Rep.* 1964;79:107–112.
19. Warnecke RB. Interventions in black populations. In: Cancer Among Black Populations. New York, NY: Alan R. Liss Inc; 1981:167–183.
20. Moss AJ, Wilder MH, NCHS. *Use of Selected Medical Procedures Associated with Preventive Care, United States: 1973.* Rockville, Md: US Dept of Health, Education and Welfare; 1977. DHEW publication HRA 77-1538.
21. Centers for Disease Control. Cancer screening behaviors among U.S. women: breast cancer, 1987–1989, and cervical cancer, 1988–1989. *MMWR.* 1992;41:17–34.
22. Calle EE, Flanders WD, Thun MJ, Martin LM. Demographic predictors of mammography and pap smear screening in US women. *Am J Public Health.* 1993;83:53–60.
23. McKenna MT, Speers M, Mallin K, Warnecke R. Agreement between patient self-reports and medical records for pap smear histories. *Am J Prev Med.* 1992;8:287–291.
24. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50212.
25. Reis LAG, Hankey BF, Miller BA, Hartman AM, Edwards BK. *Cancer Statistics, 1973–88.* Bethesda, Md: National Cancer Institute; 1991. NIH publication 91-2789.

Access to Family Planning Services and Health Insurance among Low-Income Women in Arizona

Bradford Kirkman-Liff, DrPH, and Jennie Jacobs Kronenfeld, PhD

ABSTRACT

This paper uses 1984 and 1989 Arizona survey data to explore the importance of insurance coverage for access to family planning services. In Arizona, Medicaid was provided through the Arizona Health Care Cost Containment System, a managed care model. Family planning was not provided in 1984 but had been added by 1989. Low-income women were 2.3 times more likely to receive services in 1989 than in 1984. Women in the system were 60% more likely to have received services. To improve access, health reform proposals should explicitly include family planning in managed care benefits packages. (*Am J Public Health.* 1994;84:1010–1012)

Introduction

How important is insurance coverage, including Medicaid and employment-based insurance, in ensuring access to family planning services for low-income women? These questions are addressed through analysis of 1984 and 1989 Arizona telephone survey data. During the discussion of health care reform, these are especially important questions to answer.[1,2] Many reform proposals advocate development of a minimum, basic health benefits package; some exclude family planning as a covered service.[3,4] The fate of special family planning programs such as those funded under Title XX and Title X is unclear.[5]

Reform proposals include use of managed care (e.g., health maintenance organizations, preferred provider organizations) to contain costs. Medicaid is not excluded from reform: Medicaid recipients have a choice among competing plans rather than a separate Medicaid program.[6,7] These proposals generally assume that family planning services will continue to be covered by Medicaid, but as part of a package of services through managed care.[8,9] One concern is that preventive services, such as family planning, might be underprovided to Medicaid women less skilled in seeking care under managed care systems.

This paper uses Arizona survey data from 1984 and 1989 to examine the use of family planning services among low-income women in the state's managed care Medicaid program (the Arizona Health Care Cost Containment System). Of special interest is the impact of a change of policy in this program.

Until 1981, Arizona had no Medicaid program. A managed care/health maintenance organization alternative to traditional fee-for-service Medicaid—the Arizona Health Care Cost Containment System—was begun in 1982. The program operates through a network of capitated, managed care plans with primary care physicians as gatekeepers who coordinate all services. Plans competitively bid, negotiate for contracts, and are at financial risk for overuse. Eligible recipients must obtain all care from their selected plan.[10,11] When the program was implemented, waivers excluding family planning services were granted. This waiver was eliminated in 1986, and program contractors had to either provide family planning services or pay for referrals. Most types of family planning were included (oral contraceptives, intrauterine devices, sterilization, and implants). Using survey data from 1984 (when family planning was a waived benefit) and 1989 (when family planning was covered), we examined the impact of this policy change. Because the surveys included women who were uninsured and who had employment-based insurance, we can also compare access across forms of insurance.

Methods

Survey data were collected in 1984 and 1989 as part of a larger study investigating health status and barriers to health care. Both surveys were conducted by Louis Harris and Associates Inc and commissioned by the Flinn Foundation. The 1984 survey consisted of 2508 interviews with low-income adults. The 1989 survey consisted of 3104 interviews with adults of all income levels (1847 of low income). Low income was defined as 200% of the eligibility level of the state's medically indigent component of Medicaid in that year. Rural areas were oversampled in both surveys. Samples were weighted to the latest Bureau of the Census population parameters on region, education, sex, race, and age. Further details on this survey have been published

The authors are with the School of Health Administration and Policy, College of Business, Arizona State University, Tempe.

Requests for reprints should be sent to Bradford Kirkman-Liff, DrPH, School of Health Administration and Policy, College of Business, Arizona State University, Box 874506, Tempe, AZ 85287-4506.

This paper was accepted November 12, 1993.

Note. All opinions expressed are the authors' and do not necessarily represent those of the Flinn Foundation.

elsewhere.[12] For these analyses, a subsample of low-income respondents was drawn from the 1989 survey that matched the low-income definition used in 1984, adjusted for inflation.

In addition to insurance status and the Arizona Health Care Cost Containment System, other variables were hypothesized to affect use of family planning services.[13,14] Different cultural values of Latino women may lead to different use patterns.[15] (Numbers were too small to allow analyses of African-American and Native American women.) Higher levels of educational attainment would affect receptivity to family planning messages, increasing use. Women living in rural Arizona would have lower rates of use of family planning services, given less access to clinics and greater hostility toward services.[16] Age would have a strong effect, with highest use among younger women. Employed women are more likely to use family planning services.

Univariate analysis examines percentage differences and changes over time. A more rigorous approach involves the use of logistical regression to model the dichotomous measure, simultaneously considering all independent variables. The results were transformed into odds ratios and 95% confidence intervals; odds ratios that did not include 1.0 in the confidence range can be considered statistically significant at a 95% confidence level. Analyses were conducted for the entire sample and repeated for women 18 to 29 years old.

TABLE 1—Frequency of Women Who Received Family Planning Counseling or Services

	Sample Size		Women Who Received Family Planning Counseling or Services, %		
	1984	1989	1984	1989	*P*
Subjects	997	298	17.2	33.2	.0001
Age, y					
18–29	585	178	21.0	42.0	.0001
18–20	145	32	18.6	40.8	.014
21–24	205	62	23.4	49.8	.006
25–29	235	84	20.4	35.2	.110
30–39	355	120	11.3	12.8	.800
30–34	191	67	15.7	14.4	.892
35–39	164	53	6.1	10.8	.498
Current insurance					
Uninsured	303	86	16.2	29.3	.069
Insured	694	212	17.6	34.9	.0001
AHCCCS (Medicaid)	238	113	24.4	34.2	.202
Employment-based insurance	456	99	12.6	37.4	.002
Household employment status					
At least one adult employed full-time	466	130	16.7	37.2	.001
At least one adult employed part-time and no adult employed full-time	160	42	16.9	27.5	.256
No adult employed	371	126	27.8	30.7	.069
Ethnicity					
Anglo	321	126	25.2	28.5	.617
Latino	339	127	13.0	34.4	.0001
Education					
Less than high school	450	111	17.1	31.6	.013
High school	316	98	17.7	36.0	.012
More than high school	220	88	16.8	32.5	.106
Residence					
Rural Arizona	446	112	15.0	28.2	.117
Urban Arizona	551	186	18.9	34.4	.001
Pima County (Tucson)	189	81	19.1	32.2	.135
Maricopa County (Phoenix)	362	105	18.8	35.4	.005

Note. AHCCCS = Arizona Health Care Cost Containment System.

Results

Univariate Analysis

Table 1 presents the percentage of low-income Arizona women who received family planning services in 1984 and 1989 and chi-square tests for the difference between 1984 and 1989. There was a significant increase in use of family planning services over these 5 years, from 17.2% in 1984 to 33.2% in 1989. This increase was most striking among women 18 to 20 years old; their use jumped from 18.6% in 1984 to 40.8% in 1989, with no increase for women between 30 and 39 years of age.

The shift between 1984 and 1989 varied with insurance status. Only 12.6% of low-income women with employment-based insurance had received family planning services in 1984; this significantly increased to 37.4% in 1989. Although the trends for women in Arizona's version of Medicaid and uninsured women were also in the same direction, differences were not significantly different.

Use by Anglo women did not change between 1984 and 1989, but Latino women's use significantly increased, from 13% to 34.4%. Less educated women's use also increased. There was no change for women living in rural areas; the difference for urban women was mostly due to the increase for women in the metropolitan Phoenix area.

Multivariate Analysis

As seen in Table 2, low-income women in Arizona were 2.3 times more likely to receive family planning services in 1989 than in 1984. Women between 18 and 29 years of age were also 2.3 times more likely to receive family planning services than women 30 to 39 years old. A more detailed estimation of age effects shows that women 21 to 24 years old were 4 times more likely than women 35 to 39 years old to receive family planning services. Overall, insured women were not more likely to receive family planning services when all other independent measures were considered. Women in the Arizona Health Care Cost Containment System were 60% more likely than uninsured women to receive services, with no difference for low-income women with employment-based insurance. After the multivariate approach had been used, Latino women were only 60% as likely as Anglo women to have used family plan-

TABLE 2—Multivariate Logistic Regression Results on Family Planning Use: Women 18 to 39 Years Old

	Odds Ratio	95% CI
Year		
1984	1.0	. . .
1989	2.3	1.5, 2.7
Age, y		. . .
18–29	2.3	1.7, 3.2
30–39	1.0	
18–20	3.2	1.9, 5.6
21–24	4.0	2.4, 6.5
25–29	2.7	1.6, 4.4
30–34	1.8	1.1, 3.1
35–39	1.0	. . .
Current insurance		
Uninsured	1.0	. . .
Insured	1.1	0.8, 1.4
AHCCCS (Medicaid)	1.6	1.2, 2.3
Employment-based insurance	0.8	0.5, 1.1
Household employment status		
At least one adult employed full-time	1.0	. . .
At least one adult employed part-time and no adult employed full-time	1.0	0.6, 1.5
No adult employed	1.0	0.7, 1.3
Ethnicity		
Anglo	1.0	. . .
Latino	.6	0.4, 0.8
Education		
Less than high school	1.0	0.7, 1.4
High school	1.0	. . .
More than high school	0.8	0.5, 1.2
Residence		
Rural Arizona	1.0	. . .
Urban Arizona	1.2	0.9, 1.6
Pima County (Tucson)	1.2	0.8, 1.7
Maricopa County (Phoenix)	1.2	0.9, 1.6

Note. AHCCCS = Arizona Health Care Cost Containment System; CI = confidence interval.

ning services, despite the large increase in the percentage of Latino women using family planning services between the 2 years. There was no significant relationship between household employment status, education, or residence and family planning use when all other measures were considered.

Results for women 18 to 29 years old indicated that low-income women in this age group in Arizona were 2.4 times more likely to receive family planning services in 1989 than in 1984. Overall, the model for younger women and the model for all women are very similar.

Discussion and Conclusions

The results found in this study are surprising. We expected low-income women who were insured to have greater use of family planning services than women without insurance, yet this was not the case. Only when insurance status was separated into those with Medicaid and those with employment-based insurance was there improved access to family planning services, but only for the Medicaid-covered population. One probable explanation is that most employment-based insurance does not cover family planning services, while Medicaid does. There was a significant increase in use of family planning services by low-income women between 1984 and 1989, indicating that the inclusion of family planning services by the Arizona Health Care Cost Containment System had an impact. Even in 1984, women covered by the program had higher rates of use, although the service was not covered. There are three possible explanations. First, the capitated provider networks may have provided family planning services out of their retained earnings, as allowed by program regulation. Second, plans may have provided referrals to family planning clinics operated by county public health departments, community health centers, or nonprofit organizations that filled the gap caused by exclusion of family planning under Arizona's version of Medicaid. Third, women in the program could use their limited dollars to pay for family planning services because they did not have to pay for other health care services.

These results have important implications for health reform proposals. Family planning must be explicitly included in any managed care basic benefits package created as part of health reform. Medicaid—including a managed care model—must also continue to include family planning. Although public and private nonprofit agencies can contribute to ensuring access to family planning services, access is improved when family planning is a covered service. □

Acknowledgment

This research was supported in part by the Flinn Foundation, Phoenix, Ariz.

References

1. Lundberg GD. National health care reform: an aura of inevitability is upon us. *JAMA.* 1992;265:2566–2567.
2. Sultz H. Health policy: if you don't know where you're going, any road will take you. *Am J Public Health.* 1991;81:418–420.
3. Shortell SM, McNerney WJ. Criteria and guidelines for reforming the U.S. health care system. *N Engl J Med.* 1990;322:463–467.
4. Simmons HE, Rhoades MM, Goldberg MA. Comprehensive health care reform and managed competition. *N Engl J Med.* 1992;327:1525–1528.
5. Radecki SE, Bernstein GS. Use of clinic versus private family planning care by low income women: access, cost and patient satisfaction. *Am J Public Health.* 1989;79:692–697.
6. Howell EH, Ellwood MR. Medicaid and pregnancy: issues in expanding eligibility. *Fam Plann Perspect.* 1991;23:123–128.
7. Torres A, Kenney AM. Estimating Medicaid coverage for pregnant women: estimates of the impact and cost. *Fam Plann Perspect.* 1989;21:19–24.
8. Grannemann TW. Priority setting: a sensible approach to Medicaid policy? *Inquiry.* 1991;28:300–305.
9. Merrill JC. State initiatives for the medically uninsured. *Health Care Financing Rev.* 1990;12(suppl):161–166.
10. Christianson JB, Hillman DG. *Health Care for the Indigent and Competitive Contracts: The Arizona Experience.* Ann Arbor, Mich: Health Administration Press; 1986.
11. *The Impact of Medicaid Managed Care on Public Health Systems in Arizona.* Washington, DC: Association of State and Territorial Health Officials; 1993.
12. Kirkman-Liff B, Mondragon D. Language of interview: relevance for research of southwest Hispanics. *Am J Public Health.* 1991;81:1399–1404.
13. Mosher WD, Pratt WF. Use of contraception and family planning services in the United States, 1988. *Am J Public Health.* 1990;80:1132–1133.
14. Kronenfeld JJ. *Controversial Issues in Health Care Policy.* Newbury Park, Calif: Sage; 1993.
15. Rochat R, Warren C, Smith J, Holck S, Friedman J. Family planning practices among Anglo and Hispanic women in US counties bordering Mexico. *Fam Plann Perspect.* 1981;13:176–180.
16. Firpo TH, Lewis DA. Family planning needs and services in nonmetropolitan areas. *Fam Plann Perspect.* 1976;8:231–240.

Sex Bias in the Management of Coronary Artery Disease in Quebec

William D'Hoore, PhD, Claude Sicotte, PhD, and Charles Tilquin, PhD

ABSTRACT

This study tests the hypothesis that, given the absence of financial barriers to major coronary procedures in Quebec, women are as likely as men to undergo such procedures. The use of coronary procedures in 33 940 patients with ischemic heart disease, admitted during 1 year to 78 Quebec hospitals, was analyzed. The male-to-female age- and severity-adjusted odds ratios for the use of these procedures were 1.47 for diagnostic procedures, 1.38 for therapeutic procedures, and 1.26 for diagnostic and therapeutic procedures. These results suggest that differences in the use of coronary procedures by sex are influenced by factors other than financial accessibility. (*Am J Public Health.* 1994;84:1013–1015)

Introduction

Recent studies have shown that in the United States, women with coronary artery disease are less likely than men to undergo major diagnostic and therapeutic procedures.[1,2] This puzzling finding prompted us to examine the use of coronary procedures in Quebec hospitals to assess the magnitude of a possible sex bias. As elsewhere in Canada, there is free access to care in Quebec; therefore, there is no financial barrier as there could be in the United States. Given the absence of financial barriers, our hypothesis was that women are as likely as men to undergo major coronary procedures, after adjustment for severity.

Methods

To test this hypothesis, we used the MED-ECHO hospital discharge abstract database (Quebec, Canada). The analysis included all 33 940 patients with ischemic heart disease (*International Classification of Diseases*, ninth revision[3] [ICD-9] codes 410, 411, 413, 414) admitted during the administrative year 1989/90 to the 78 Quebec hospitals with 100 or more acute-care beds. A distinction was drawn between diagnostic procedures (coronary angiography and cardiac catheterization) and therapeutic procedures used for revascularization (percutaneous transluminal coronary angioplasty and coronary artery bypass graft). Logistic regression analyses were performed to relate the use of coronary procedures to predictors including sex, age (<50, 51–60, 61–70, 71–80, and >80 years), diagnosis (myocardial infarction vs other ischemic heart disease), and the Charlson index of comorbidity.[4] The Charlson comorbidity index ranges from 0 to 24, a higher score indicating a heavier burden of comorbid disease.

Results

Summary statistics are presented in Table 1. Of the 33 940 patients, 66.32% were male. The women were older than the men and scored higher on the comorbidity index. A total of 7935 patients (23.4%) had coronary angiography or cardiac catheterization, and 5983 patients (17.6%) underwent percutaneous transluminal coronary angioplasty or coronary artery bypass graft. Both diagnostic and therapeutic procedures were performed at the same time in 1850 patients (5.5%). Age and comorbidity were inversely associated with the presence of either diagnostic or therapeutic procedures. The presence of myocardial infarction strongly reduced the likelihood of diagnostic and therapeutic procedures.

Tables 2 and 3 show, respectively, crude and case mix–adjusted relative risks for the use of coronary procedures. The bivariate and multivariate estimations are similar: women were less likely than men to be treated aggressively. However, once a diagnostic procedure was performed, there was no difference between men and women in the use of revascularization. Therefore, we tried to refine the analysis of the use of diagnostic procedures. Logistic models including interaction terms between sex and other variables and terms testing for nonlinearity in the logit of the dependent variable were evaluated. The interaction between age and comorbidity was found to be the only significant term. The effect of age appeared to be nonlinear: older patients were much less likely to undergo diagnostic procedures than were younger patients (for patients aged 61–70, 71–80, and >80 years in comparison with patients younger than 50 years, the relative odds were, respectively, 1, 0.60, and 0.11). In all models, the association between sex and the use of diagnostic procedures remained strongly significant.

Finally, we assessed the outcomes of care in relation to the use of coronary

William D'Hoore is with the Department of Hospital Administration, Université Catholique de Louvain, Brussels, Belgium. Claude Sicotte and Charles Tilquin are with the Département d'administration de la santé, Université de Montréal, Montréal, Québec, Canada.

Requests for reprints should be sent to Claude Sicotte, PhD, Université de Montréal, Département d'administration de la santé, C.P. 6128, Succursale A, Montréal, Québec, Canada H3C 3J7.

This paper was accepted September 16, 1993.

TABLE 1—Characteristics of Patients with Ischemic Heart Disease, Quebec, 1989/90

	Men (n=22508)	Women (n=11432)
Diagnosis,[a] %		
410	34.37	35.49
411	10.47	12.74
413	24.18	27.94
414	30.98	23.84
Age, %		
<50 y	20.24	7.41
51–60 y	26.39	15.36
61–70 y	31.39	30.90
71–80 y	16.95	30.62
>80 y	5.03	15.71
Charlson comorbidity index level, %		
0	62.63	51.44
1–2	27.63	33.15
3–4	7.21	11.83
5–6	1.78	2.43
>6	0.75	1.15
Alive, %	94.12	89.45

[a]*International Classification of Diseases*, Ninth Revision, codes.[3]

TABLE 2—Unadjusted Male-to-Female Relative Risks for the Use of Coronary Procedures

	Relative Risk	95% Confidence Interval
Angiography	1.66	1.55, 1.79
Catheterization	1.51	1.43, 1.60
Angiography or catheterization	1.58	1.51, 1.65
Revascularization	1.53	1.45, 1.61
Diagnostic and therapeutic procedures	1.48	1.34, 1.64

TABLE 3—Adjusted Male-to-Female Odds Ratios for the Use of Coronary Procedures

	Odds Ratio	95% Confidence Interval
Catheterization or angiography	1.47	1.38, 1.57
Revascularization	1.38	1.29, 1.48
Revascularization once a diagnostic procedure has been performed	0.93	0.82, 1.06
Diagnostic and therapeutic procedures	1.26	1.13, 1.41
Percutaneous transluminal coronary angioplasty (vs coronary artery bypass graft)	0.69	0.61, 0.78

TABLE 4—Adjusted Male-to-Female Odds Ratios for Death from Ischemic Heart Disease

	n	Odds Ratio	95% Confidence Interval
Total sample	33 940	0.89	0.81, 0.97
With revascularization	5 983	0.81	0.55, 1.20
With no revascularization	27 957	0.89	0.81, 0.98

procedures (Table 4). The overall mortality of women was higher than the mortality of men, but this difference disappeared when only patients who underwent a therapeutic procedure were considered. On this subject, we must note that the accuracy of fit of logistic models, measured by the area under the receiver operating characteristic curve,[5,6] was higher for the prediction of death (about 0.86, a level as good as can be reached by models with clinical data[7]) than for the prediction of use of coronary procedures (0.67 or less).

Analyses performed separately on low-risk patients (age < 70 years and comorbidity index = 0) and high-risk patients (age > 70 years and comorbidity index > 1) led to the same results: women with coronary disease were less likely than men to undergo coronary procedures.

Discussion

The differences between men and women in the use of coronary procedures are striking. Our results show that physicians' decisions to perform diagnostic procedures vary nonlinearly according to the age of patients, as if an implicit cutoff were present. Patients' preferences may explain this result. For example, older patients with ischemic heart disease are possibly more often widowed or alone, a factor that may deter them from seeking aggressive care.[8,9] However, age does not completely account for the differences between the sexes. Physicians may believe that diagnostic procedures are more risky in women and that the subsequent benefit of performing revascularization is smaller for women than for men. Physicians' decisions seem to depend on the patient's sex instead of his or her clinical status. However, given the poor fit of logistic models that predict procedure use, the clinical context that allows physicians to appropriately prescribe coronary procedures may not be grasped by such variables as age and comorbidity.

With regard to the higher mortality of women, it seems that although it is less frequent, ischemic heart disease is more severe in women than in men. Women may be referred for coronary surgery at a more advanced stage of disease than are men.[10]

We do not know whether coronary procedures are underused in women and overused in men, because it was not possible to assess the appropriateness of these procedures. Therefore, physicians should readjust the management of coronary disease in both women, who may not be undergoing beneficial procedures, and men, who may be undergoing procedures that are not beneficial. However, we are inclined to favor the hypothesis that coronary procedures are underused in women, given women's higher mortality, which is interpreted here as a referral bias.

Since our results are nearly identical to what was found in the United States,[1,2] they suggest that sex differences in the use of coronary procedures are not influenced by factors such as health services organization or financial accessibility. □

References

1. Ayanian JZ, Epstein AM. Differences in the use of procedures between women and men hospitalized for coronary heart disease. *N Engl J Med.* 1991;325:221–225.
2. Steingart RM, Packer M, Hamm P, et al. Sex differences in the management of coronary artery disease. *N Engl J Med.* 1991;325:226–230.
3. *Manual of the International Classification of Diseases, Injuries, and Cause of Death, 1975, 9th Revision.* Geneva, Switzerland: World Health Organization; 1977.
4. Charlson ME, Pompei P, Ales KL, McKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40:373–383.
5. Erdreich LS, Lee ET. Use of relative operating characteristic analysis in epidemiology. *Am J Epidemiol.* 1981;114:649–662.
6. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143:29–36.
7. Higging TL, Estafanous FG, Loop FD, Beck GJ, Blum JM, Paranandi L. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients—a clinical severity score. *JAMA.* 1992;267:2344–2348.
8. Greenberg ER, Chute CG, Stukel T, et al. Social and economic factors in the choice of lung cancer therapy. *N Engl J Med.* 1988;318:612–617.
9. Goodwin JS, Hunt WC, Key CR, Samet JM. The effect of marital status on stage, treatment, and survival of cancer patients. *JAMA.* 1987;258:3125–3130.
10. Kahn SS, Nessim S, Gray R, Czer LS, Chaud A, Matloff J. Increased mortality of women in coronary artery bypass surgery: evidence for referral bias. *Ann Intern Med.* 1990;112:561–567.

Health Insurance Coverage among Persons with AIDS: Results from a Multistate Surveillance Project

Theresa Diaz, MD, MPH, Susan Y. Chu, PhD, Lisa Conti, DVM, Bernard L. Nahlen, MD, Bruce Whyte, MD, Eve Mokotoff, MPH, Anne Shields, MHA, Patricia J. Checko, MPH, Mary Herr, MT, Quaiser Mukhtar, PhD, Cornelis A. Rietmeijer, MD, MSPH, Anna Levy, MPH, Pat Hermann, MD, and James W. Buehler, MD

ABSTRACT

To determine factors associated with health insurance coverage among persons with acquired immunodeficiency syndrome (AIDS), we interviewed 1958 persons 18 years of age or older who were reported to have AIDS in 11 states and cities. Overall, 25% had no insurance, 55% had public insurance, and 20% had private insurance. Factors associated with lack of insurance varied by current employment status. Employed persons with an annual household income of less than $10 000 were 3.6 times more likely to lack insurance than employed persons with a higher income. Unemployed persons diagnosed with AIDS for less than 1 year were two times more likely to lack health insurance than unemployed persons diagnosed for a longer time. Making insurance available to persons identified as most likely to lack insurance should improve access to care for persons with AIDS. (*Am J Public Health.* 1994;84: 1015–1018)

Introduction

Access to and use of health care services are affected by insurance status.[1–5] Increasingly, coverage for acquired immunodeficiency syndrome (AIDS) cases is shifting to public insurers, such as Medicaid,[6] most likely resulting from loss of private insurance among persons with AIDS[7–9] and demographic shifts in populations affected by the human immunodeficiency virus (HIV) epidemic.[10,11] To describe insurance coverage among persons with AIDS, we used information from a multistate interview project.

Methods

Eleven state and local health departments are collaborating with the Centers for Disease Control and Prevention (CDC) in the Supplement to HIV/AIDS Surveillance project. This project expands information routinely collected via AIDS surveillance through personal interviews of persons with AIDS. Persons reported with AIDS who are 18 years of age or older and are medically able to complete

Theresa Diaz, Susan Y. Chu, and James W. Buehler are with the Division of HIV/AIDS, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Ga. Lisa Conti is with the Florida Department of Health and Rehabilitative Services, Tallahassee, Fla. Bernard L. Nahlen is with the Division of HIV/AIDS, National Center for Infectious Diseases, and the Los Angeles County Health Department, Los Angeles, Calif. Bruce Whyte is with the Georgia Department of Human Resources, Atlanta, Ga. Eve Mokotoff is with the Michigan Department of Public Health, Detroit, Mich. Anne Shields is with the Washington Department of Health, Seattle, Wash. Patricia J. Checko is with the Connecticut State Department of Health Services, Hartford, Conn. Mary Herr is with the Division of Public Health, Delaware Health and Social Services, Wilmington, Del. Quaiser Mukhtar is with the Arizona Department of Health, Phoenix, Ariz. Cornelis A. Rietmeijer is with the Denver Department of Health and Hospitals, Denver, Colo. Anna Levy is with the New Mexico Health Department, Albuquerque, NM. Pat Hermann is with the South Carolina Department of Health and Environmental Control, Columbia, SC.

Requests for reprints should be sent to Theresa Diaz, MD, MPH, Centers for Disease Control and Prevention, Division of HIV/AIDS, 1600 Clifton Rd, MS E47, Atlanta, GA 30333.

This paper was accepted November 17, 1993.

TABLE 1—Characteristics of Respondents and Reported AIDS Cases among Persons 18 Years of Age or Older, in Project Areas and Nationally, June 1, 1990, through June 30, 1992

Characteristic	Respondents in Project Areas, %		
	Interviewed (n = 1958)	Not Interviewed (n = 717)	Persons with AIDS in the United States, % (n = 92 403)[a]
Male gender*	85	92	87
Male-male sex exposure group*	59	58	54
Injection drug user	19	17	24
Male-male sex and injection drug user	8	11	5
Heterosexual contact	11	5	8
Transfusion associated	1	2	3
Undetermined	2	7	6
Race/ethnicity*			
White	52	55	49
Black	35	29	32
Hispanic	10	12	18
Other[b]	3	4	1

[a]Includes persons reported with AIDS during the same time period respondents were interviewed (June 1, 1990 through June 30, 1992).
[b]Includes persons of Asian, Pacific Islander, American Indian, and Alaskan native race/ethnicity.
*$P < .05$, interviewed compared with not interviewed.

the interview are eligible, as previously described.[12] Six sites used a population-based sample in which interviews were attempted with all eligible persons (Arizona, Delaware, South Carolina, and New Mexico) or with a 30% random sample of men who have sex with men (Los Angeles and Washington) and all persons with other modes of transmission. Five sites (Denver, Detroit, Connecticut, Florida, and Georgia) used a facility-based sample in which interviews were attempted with all eligible persons at selected medical facilities. Because these facilities were mostly public hospitals or clinics, insurance coverage varied by sampling method. Each participating health department developed procedures to ensure patient confidentiality, and projects were approved by local human subjects review boards.

We analyzed data for persons reported to have AIDS from June 1, 1990, through June 30, 1992. Participants were asked whether they currently had health insurance (including public insurance) and, if so, what type of insurance they mainly used to pay for HIV-related health care. For purposes of analysis, private insurance included both self-paid insurance (including self-paid COBRA insurance) and employer-provided insurance; public insurance included Medicaid, Medicare, state-funded insurance continuation (including state-funded COBRA insurance), the Civilian Health and Medical Program of the Uniformed Services, and coverage by the Veterans Administration. (The Consolidated Omnibus Budget Reconciliation Act [COBRA] of 1985 allows individuals to continue private health insurance they received when employed for 18 to 29 months after loss of employment. States can pay health insurance premiums, with matching federal funds, for individuals who stopped working because of illness and can no longer afford premiums.) Income was defined as household income in the year preceding interview, before taxes.

In a multivariate analysis, logistic regression[13] was used to determine factors associated with no insurance. We included variables significantly associated with no insurance in the univariate analysis (time since AIDS diagnosis and sampling method), as well as two variables most likely to change over time (employment and income). Because factors associated with lack of insurance differed by employment status, odds ratios were calculated separately for unemployed and employed persons. When results differed by sampling method, odds ratios were calculated separately.

Results

Description of Respondents

Of the 2675 eligible persons with AIDS selected for interviews, 1958 (74%) completed interviews, 331 (12%) refused to be interviewed, and 386 (14%) could not be located. Compared with persons not interviewed, those who were interviewed were more likely to be female, to have a known exposure mode, and to be Black (Table 1). Characteristics of persons interviewed were comparable to those of persons with AIDS reported in the United States during the study interval except for a lower proportion of Hispanics (Table 1). The median age of respondents was 36 years (range = 18–77 years); 74% were 30–49 years of age.

Health Insurance Coverage

Overall, 25% of respondents had no insurance, 20% had private insurance, and 55% had public insurance (Table 2). Among those with private insurance, 71% had employer-provided insurance, and 29% had self-paid insurance. Among persons with public insurance, 90% had Medicaid, 4% had Medicare, 3% had state-funded assistance programs, 2% were part of the Civilian Health and Medical Program of the Uniformed Services or the Veterans Administration, and 1% had state-funded insurance continuation (including COBRA).

Eighty percent of persons with public insurance and 51% of persons with private insurance stated that their insurance paid for all treatments. Respondents with Medicaid had the highest proportion (83%) stating that all services were covered, and persons with Medicare had the lowest proportion (18%). Forty-eight percent of persons with private insurance and 74% of persons with Medicare said their insurance paid for some but not all treatments.

In the univariate analysis, characteristics significantly ($P < .05$) associated with not having insurance included being diagnosed with AIDS for less than 1 year and being selected from a facility-based sample (Table 2). In the multivariate analysis, among employed persons, having a household income of less than $10 000 in the past year (adjusted odds ratio [AOR] = 3.6, 95% confidence interval [CI] = 2.6, 5.1) was associated with lack of health insurance; no association was found with time since AIDS diagnosis. Among unemployed persons from the population-based sample, having an income of less than $10 000 (AOR = 1.7, CI = 1.2, 2.5) and being diagnosed with AIDS for less than 1 year (AOR = 2.2, CI = 1.4, 3.4) were associated with lack of health insurance. Conversely, among unemployed persons from facility-based samples, per-

sons with an income of less than $10 000 (AOR = 1.7, CI = 1.2, 2.0) were more likely to have insurance than persons with a higher income; being diagnosed with AIDS for less than 1 year continued to be associated with no insurance (AOR = 2.1, CI = 1.7, 2.6).

Discussion

Among persons with AIDS, eligibility for public insurance has increased and public resources are used extensively for health care.[5,6,14,15] Regardless of race/ethnicity, sex, or HIV exposure mode, we found the majority of our respondents relied on public insurance (mostly Medicaid) for medical care. Yet despite the increased eligibility for publicly funded insurance for persons with AIDS,[16,17] one quarter of our respondents still lacked health insurance, similar to the proportion of all persons in the United States age 18–44 years who lack health insurance (14%–27%).[18,19] A preliminary analysis of 821 persons with HIV infection but not AIDS who were interviewed as part of this project found that an even higher proportion (40%) reported no health insurance (CDC, unpublished data).

In the United States, income is an important predictor of lack of health insurance coverage, with 47% of those with incomes below the poverty level lacking health insurance in 1987.[19,20] We found this generalization to be true for all employed persons and for unemployed persons from the population-based sample. However, among the unemployed in the facility-based sample, persons with lower incomes were more likely to have insurance. This finding is not surprising because although overall the facility-based sample had a higher proportion of persons with no insurance (as would be expected in public medical facilities), persons attending these facilities are probably provided with assistance in acquiring public insurance, and unemployed persons with low incomes would likely qualify. Regardless of sampling method, unemployed persons who had been diagnosed with AIDS for less than 1 year were more likely to lack health insurance than unemployed persons who had been diagnosed for at least 1 year. This difference may reflect ineligibility for Medicaid because personal assets were too high or because of administrative delays in obtaining public insurance or time restrictions for Medicare eligibility.[15,20] For example, persons who were once employed and now receive Social Security Disability Income may have an income too high to qualify for Medicaid, but must wait 2 years before they have the option to buy Medicare.[16,17,21] Unless they receive funds under COBRA, such recently unemployed persons are unlikely to be able to continue the private health insurance they had when employed.[17,20]

Even among insured persons, inequities in health care may occur. We found differences in payments for medical treatment for HIV infection by insurance type, with Medicare recipients being the least likely to report that all services were covered by their insurance. The need to supplement insurance coverage with out-of-pocket funds may deter appropriate use of health care and result in adverse clinical outcomes.[3]

Improving the use of, and access to, health services for HIV-infected persons will require increased insurance coverage for both employed and unemployed persons. □

TABLE 2—Proportion of Persons within Each Group with Private, Public, or No Health Insurance

Characteristic	Sample Size	Insurance Type, %		
		Private	Public	None
Sex				
Male	1632	21	53	26
Female	284	11	65	24
Race/ethnicity				
White	965	26	50	24
Black	679	11	62	27
Latin American	122	25	46	29
Puerto Rican	66	9	64	27
Exposure group				
Male-male sex	1122	26	50	24
Injection drug use	367	8	63	28
Male-male sex and injection drug use	143	15	57	28
Heterosexual contact	218	12	61	27
Education				
≥ 12 y	1399	25	50	25
< 12 y	501	6	67	27
Annual household income				
< $10 000	1083	4	69	27
≥ $10 000	746	43	33	24
Current employment				
Unemployed	1525	10	64	26
Employed	390	60	14	26
Time since AIDS diagnosis				
< 1 y	1317	22	49	29
≥ 1 y	599	16	67	17*
Sampling				
Facility based	1311	11	59	30
Population based	605	40	45	15*
Total	1958	20	55	25

Note: Excluded are 42 (2.1%) persons with unknown insurance status; respondents for whom there was no information were excluded within each group.
*Significant difference in the proportion with no insurance ($P < .05$); the largest group in each category is the reference group (e.g., facility-based samples have a significantly higher proportion of persons with no insurance compared with population-based samples).

Acknowledgments

We would like to thank the many interviewers in participating health departments whose efforts made this project possible. We would also like to thank Bob Byers, PhD, for his statistical consultation; W. Meade Morgan, PhD, Carolyn Sanchez, and Wyndy Amerson for software development and support; Linda Leary for data management and general support; and Laurence Slutsker, MD, Russell Gerber, MD, and Penelope Pine, PhD, for their review of the manuscript.

References

1. Anderson R, Newman J. Societal and individual determinants of medical care utilization in the United States. *Milbank Mem Fund Q.* 1973;51:95–124.
2. Hayward RA, Shapiro MF, Freeman HE, Corey CR. Inequalities in health services among insured Americans. Do working-

age adults have less access to medical care than the elderly? *N Engl J Med.* 1988;318: 1507–1512.
3. Bodenheimer T. Underinsurance in America. *N Engl J Med.* 1992;327:274–278.
4. Loue S, Fullerton SC, Mathews WC. The effects of patient medical insurance on survival in a cohort of HIV infected patients. *Clin Res.* 1993;41:290A.
5. Fleishman JA, Mor V. Insurance status among people with AIDS: relationships with sociodemographic characteristics and service use. *Inquiry.* 1993;30:180–188.
6. Green J, Arno PS. The "Medicaidization" of AIDS. Trends in the financing of HIV-related medical care. *JAMA.* 1990;264: 1261–1266.
7. Yelin EH, Greenblatt RM, Hollander H, McMaster JR. The impact of HIV-related illness on employment. *Am J Public Health.* 1991;81:79–84.
8. Kass N, Munoz A, Chen B, et al. Change in employment, insurance, and income in relation to HIV status and disease progression. In: Proceedings from the VIII International Conference on AIDS; 1992; Amsterdam, the Netherlands. Abstract PoD 5579 (p D483).
9. Kass NE, Faden RR, Fox R, Dudley J. Loss of private health insurance among homosexual men with AIDS. *Inquiry.* 1991; 28:249–254.
10. Centers for Disease Control and Prevention. Update: acquired immunodeficiency syndrome–United States, 1992. *MMWR Morb Mortal Wkly Rep.* 1993;42:547–557.
11. Diaz T, Buehler JW, Castro KG, Ward JW. AIDS trends among Hispanics in the United States. *Am J Public Health.* 1993;83: 504–509.
12. Diaz T, Chu SY, Frederick M, et al. Sociodemographics and HIV risk behaviors of bisexual men with AIDS: results from a multistate interview project. *AIDS.* 1993;7:1227–1232.
13. SAS Institute Inc. *SAS/STAT Users Guide, Version 6.* 4th ed. Cary, NC: SAS Institute Inc; 1989;1:1077–1078.
14. Baily MA, Bilheimer L, Wooldridge J, Langwell K, Greenber W. Economic consequences for Medicaid of human immunodeficiency virus infection. *Health Care Financing Rev.* 1990 (annual supplement):97–108.
15. Hannan J. *America Living with AIDS.* Report of the National Commission on Acquired Immune Deficiency Syndrome, 1991. Washington, DC: US Government Printing Office; 1991.
16. AIDS patients eligible for benefits under new social security regulations. *Public Health Rep.* 1988;103:327.
17. *Financing Health Care for People with AIDS: The Role of the Health Care Financing Administration.* Washington, DC: Health Care Financing Administration; 1990.
18. Freidman E. The uninsured. From dilemma to crisis. *JAMA.* 1991;265:2491–2495.
19. *Money Income of Households, Families, and Persons in the United States: 1991.* Washington, DC: US Bureau of the Census; 1992. Current population reports, series P-60, No. 180.
20. Ellwood MR, Fanning TR, Dodds S. Medicaid eligibility patterns for persons with AIDS in California and New York, 1982–1987. *J Acquir Immune Defic Syndr.* 1991;4:1036–1045.
21. Pascal A, Cvitanic M, Bennett C, Gorman M, Serrato CA. State policies and the financing of acquired immunodeficiency syndrome care. *Health Care Financing Rev.* 1989;11:91–104.

The Reliability of Racial Classifications in Hospital Discharge Abstract Data

Jan Blustein, MD, PhD

ABSTRACT

Many studies demonstrating an association between race and the use of medical services have used hospital discharge abstract data. The quality of the measures of race in such data sources has heretofore been unexplored. Hospital discharge abstract data from New York State were used to identify 767 cardiac patients who had been admitted to a hospital twice. Racial classifications during the two admissions were concordant 93.7% of the time. Kappa was .89 for Blacks, .72 for Whites, and .43 for all other racial groups. Evidence suggests that the misclassification of race in hospital discharge abstract data is nondifferential; racial discrepancies in access to medical services are thus probably even greater than those previously reported. (*Am J Public Health.* 1994;84: 1018–1021)

Introduction

Numerous studies have shown that Black Americans receive fewer medical services than their White counterparts.[1-8] Much of that research has been based on data abstracted from hospital medical records and reported to administrative agencies. However, there have been no published investigations of the reliability or validity of patient racial classifications in hospital discharge data sets. This is remarkable for two reasons. First, several studies have cast doubt on the reliability of the clinical information in those data sources, raising broader questions about the overall quality of the measures therein.[9-11] Second, critics have increasingly questioned the validity of the racial classifications in secondary data sources.[12]

This report of the reliability of racial classifications in a hospital discharge data source examines the concordance between assigned classifications during successive admissions for patients admitted twice. It describes the likely implications of misclassification for bias in previous reports of interracial differences in service use. Finally, it describes some difficulties that arise in validating measures of race.

Methods

Source of the Data

The data source for the study was the Statewide Planning and Research Cooperative System (SPARCS), a hospital discharge abstract database maintained by the New York State Department of Health. It includes demographic, clinical, and billing information and is based on reports from individual hospitals. For each patient who is discharged, hospitals cull information from various sources. For example, clinical information is ab-

The author is with the College of Physicians and Surgeons and the School of Public Health, Columbia University, New York, NY.

Requests for reprints should be sent to Jan Blustein, MD, PhD, Division of General Medicine, College of Physicians and Surgeons, Columbia University, 630 W 168th St, Room 875 PH, New York, NY 10032.

This paper was accepted February 15, 1994.

stracted from patient medical records by trained record abstractors.

Interestingly, demographic classifications such as gender and race are not based on information abstracted from medical records. Rather, according to various knowledgeable sources, they are based on entries recorded by admitting clerks at the time that patients enter the hospital (oral communication, R. Davis, New York State Department of Health, September 1992; oral communication, B. Pavelcheck, Hospital Association of New York State, February 1994). According to those sources, race is generally assigned on the basis of encounters with patients or their proxies. There are no formal rules for assigning race to a patient. Because of the sensitive nature of racial identity, however, classification is often made by clerks' observations rather than by direct questioning.

Such procedures are apparently common in other states in which hospital discharge data are collected. For example, according to a document released by the California State Department of Health, racial classifications recorded by clerical staff at the time of admission are also generally the source of the race variable in that state's hospital discharge abstract system.[13] However, "classification of race/ethnicity is a difficult task. Hospital personnel [are] often unsure how to classify patients . . . [but are] reluctant to ask the patient directly."[13] In those cases, admitting clerks assign race as best they can.

In New York State, demographic, clinical, and billing data from each admission are merged to create the electronic database. For individuals admitted to a hospital more than once, records of those admissions can be linked by matching patient insurance billing numbers.[14] In this study, discharges were linked (i.e., ascribed to the same individual) if they had the identical billing number and reflected the same gender and date of birth.

Subjects

The 9576 subjects were part of a prior study of patients admitted to the hospital with a principal diagnosis of myocardial infarction in 1986.[15,16] During that index admission, there were substantial racial differences in cardiac service use (J. Blustein, unpublished data, June 1993). Specifically, patients classified as White were more than 60% more likely to receive a high-technology cardiac service (cardiac catheterization, angioplasty, or bypass surgery) than were patients classified as Black (odds ratio [OR] = 1.69, 95% confidence interval [CI] = 1.14, 2.50).

For various reasons, some of the patients were readmitted to a hospital during the subsequent 6-month period (hereinafter termed a "subsequent admission"). Subjects were selected for further analysis if they had been readmitted to a hospital other than the index hospital during a subsequent admission. Of the 9576 patients, 2487 were subsequently admitted; of those individuals, 767 were readmitted to a different hospital. None were admitted to three different hospitals.

Classification of Race

Race is classified as Asian, Black, American Indian/Eskimo/Aleut, "other," and White in the Statewide Planning and Research Cooperative System data set, consistent with federal guidelines.[17] A small number of patients are listed as "unknown." Because researchers using hospital discharge data have typically combined Asian, American Indian/Eskimo/Aleut, other, and unknown into the category "all other," the classifications Black, White, and all other were used for some portions of the analyses reported here.

Analytic Approach

Standard measures of crude concordance and kappa, the interrater reliability coefficient,[18] were used to calculate the extent of agreement between racial classifications during the index and subsequent admissions.

Results

In Table 1, racial classifications of the 767 subjects who were admitted to two different hospitals are compared with those of patients who were not admitted to two different hospitals. The two groups did not differ with respect to racial classification during the index admission ($\chi^2 = 6.2, 5\ df, P = .28$).

As shown in Table 2, racial classifications during the index and subsequent admissions were concordant in 719 of the 767 subjects (93.7%). The resultant percentage agreement and kappa for each racial classification are presented in Table 3. The reliability coefficients were .89 (95% CI = .82, .96) for Black race and .72 (95% CI = .64, .80) for White race. For other categories, reliability was even less consistent. When the other categories (other, Asian, American Indian, unknown) were combined, the interrater reliability coefficient was .43 (95% CI = .28, .58). Assessment of the reliability of the individual non-Black, non-White categories was limited by small sample sizes.

TABLE 1—Racial Classification of Patients during Their Index Admissions, by Whether They were Admitted to Two Different Hospitals (n = 9576)

	Admitted to Two Different Hospitals	
	Yes (n = 767), % (No.)	No (n = 8809), % (No.)
Black	6.4 (49)	6.3 (556)
White	89.0 (683)	88.6 (7806)
Other	3.9 (30)	4.1 (364)
Asian	0.3 (2)	0.5 (47)
American Indian	0.3 (2)	0.1 (5)
Unknown	0.1 (1)	0.4 (41)

Finally, as shown in Table 4, the use of cardiac services during the index admission was not associated with concordance between racial classifications during the index and subsequent admissions ($\chi^2 = 1.48, 1\ df, P = .22$).

Discussion

This study of hospital discharge abstract data revealed a substantial lack of reliability for racial classifications, particularly for non-Black, non-White categories. This pattern of results is consistent with previous reports comparing birth and death records[19] and with studies of discrepancies between self- and observer-classified race.[20]

As described above, the data analyzed here emanate from a study in which White cardiac patients were estimated to be 69% more likely than their Black counterparts to receive a high-technology cardiac service. The present findings suggest that this estimate of the effect of race may have been biased by measurement error. Is it possible to assess the extent of that bias? More generally, are estimates of the impact of race based on hospital discharge data sets likely to overstate or understate true racial disparities in access to services?

If misclassification of patient race is nondifferential (i.e., if the tendency to

TABLE 2—Racial Classification during Index and Subsequent Admissions for Patients Admitted to Two Different Hospitals (n = 767)

Classification during Index Admission	Classification during Subsequent Admission						
	Black	White	Other	Asian	American Indian	Unknown	Total
Black	41	5	3	0	0	0	49
White	0	665	17	1	0	0	683
Other	1	15	11	2	0	1	30
Asian	0	1	0	1	0	0	2
American Indian	0	1	0	0	1	0	2
Unknown	0	0	1	0	0	0	1
Total	42	687	32	4	1	1	767

TABLE 3—Percentage of Concordance and Interrater Reliability (Kappa) of Reported Racial Classifications during Index and Subsequent Admissions (n = 767)

Racial Classification	Concordance, %	Kappa	95% Confidence Interval
Black (n = 49)	99	.89	.82, .96
White (n = 683)	95	.72	.64, .80
All other (n = 35)	95	.43	.28, .58
Other (n = 30)	95	.32	.17, .47
Asian (n = 2)	99	.33	−.16, .82
American Indian (n = 2)	99	.66	0.4, 1.00
Unknown (n = 1)	99	−.00	.00, .00

Note. Percentage of concordance refers to the proportion of the 767 cases in which the two successive classifications were concordant for the dichotomy in question. For example, in 99% of the cases, patients were concordantly classified according to the dichotomy Black/non-Black. "All other" is an aggregated classification, including patients classified as other, Asian, American Indian, and unknown (see text).

TABLE 4—Concordance in Racial Classification and Rate of Use of Cardiac Services for Patients Admitted to Two Different Hospitals (n = 767)

	Race Classified Concordantly	
	Yes (n=719), % (No.)	No (n=48), % (No.)
Received cardiac services	12.1 (87)	6.3 (3)
Did not receive cardiac services	87.9 (632)	93.8 (45)

misclassify a patient as White or Black is independent of the likelihood that that patient will receive services), then the estimated racial discrepancy will be biased toward the null value.[21] Real discrepancies in service use will be greater than those reported.

There are several reasons to believe that misclassification of patient race is non-differential with respect to service use. First, racial classifications are assigned by hospital clerks at the time of admission, and, in most cases, decisions about providing services are made after admission. In those cases, misclassification is necessarily nondifferential. It is noteworthy that patients admitted emergently—including most patients admitted with myocardial infarction—will meet this description. In other cases, patients are electively admitted in order to receive services. Although those patients (or their proxies) may announce the intended treatment to the admitting clerk, it seems rather unlikely that clerks would classify patient race on that basis. Finally, the results reported in Table 4 are consistent with misclassification being nondifferential.

Recently, several investigators have used validation studies to "correct" effect estimates based on risk factors measured with known error.[22-24] However, it is not clear how these methods could be useful in correcting reported effects of race. Determining the sensitivity and specificity of racial classifications would require that there be some "gold standard" for race. As has been emphasized by many commentators, however, no such standard exists. Race is not a biological construct; rather, racial classifications are "based on socially defined phenotypic traits as seen through the filter of individual and social perspective."[25]

In explaining racial differences in cardiac service use, commentators have proposed two competing theories. One theory is that White patients may "prefer" more services than do Black patients.[1,2,4,8] The other is that physicians' treatment decisions are influenced—either consciously or unconsciously—by patient race.[1,2,4,8] These competing theories suggest two different validation sources. If the first theory is correct, then patients' self-categorization of race would probably be the logical gold standard. On the other hand, if the second theory is true, then physicians' perceptions of patients' race would be more appropriate.

In spring 1993, the working group from a conference convened by the Centers for Disease Control and Prevention recommended that, in all public health surveillance data sources, "race and ethnicity status should be self-identified using a multiple-choice option."[12] Implementation of this recommendation might well enhance the uniformity of the data collected by various agencies. However, if the second of the above-cited theories is true, access to services may depend less on self-categorized race than on the racial classifications and perceptions of others.

In conclusion, categorization of race in hospital discharge data exhibits a pattern of measurement error similar to that previously reported in other data sources used for public health surveillance. This misclassification is probably nondifferential; thus prior reports of the impact of race on service use are probably biased downward. However, because race is subjectively determined, it is not clear how to correct these misestimates of the impact of race on the use of medical services. □

Acknowledgments

This study was supported by a grant from the United Hospital Fund and the Agency for

Health Care Policy and Research (postdoctoral fellowship 5F32HS0034).

I am grateful to Dr Steven Shea of the College of Physicians and Surgeons and Dr Daniel Rabinowitz of the Department of Statistics, Columbia University, for their helpful comments.

References

1. Wenneker MB, Epstein AM. Racial inequalities in the use of procedures for patients with ischemic heart disease in Massachusetts. *JAMA*. 1989;261:253–257.
2. Hannan EL, Kilburn H, O'Donnell JF, Lukacik G, Shields EP. Interracial access to selected cardiac procedures for patients hospitalized with coronary artery disease in New York State. *Med Care*. 1991;29:430–441.
3. Johnson PA, Lee TH, Cook EF, Rouan GW, Goldman L. Effect of race on the presentation and management of patients with acute chest pain. *Ann Intern Med*. 1993;118:593–601.
4. Maynard C, Litwin PE, Martin JS, et al. Characteristics of black patients admitted to coronary care units in metropolitan Seattle: results from the Myocardial Infarction Triage and Intervention Registry (MITI). *Am J Cardiol*. 1991;67:18–23.
5. Golberg KC, Hartz AJ, Jacobson SJ, Krakauer H, Rimm AA. Racial and community factors influencing coronary artery bypass graft surgery rates for all 1986 Medicare patients. *JAMA*. 1992;267:1473–1477.
6. Bombadier C, Fuchs VR, Lillard LA, et al. Socioeconomic factors affecting the utilization of surgical operations. *N Engl J Med*. 1977;297:699–705.
7. Gittelsohn AM, Halpern J, Sanchez RL. Income, race, and surgery in Maryland. *Am J Public Health*. 1991;81:1435–1441.
8. Whittle J, Conigliaro J, Good CB, Lofgren RP. Racial differences in the use of invasive cardiovascular procedures in the Department of Veterans Affairs medical system. *N Engl J Med*. 1993;329:621–627.
9. Fisher ES, Whaley FS, Krushat M, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health*. 1992;82:243–248.
10. Green J, Wintfel N. How accurate are hospital discharge data for evaluating effectiveness of care? *Med Care*. 1993;31:719–731.
11. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. *Ann Intern Med*. 1993;119:844–850.
12. Centers for Disease Control and Prevention. Use of race and ethnicity in public health surveillance. Summary of the CDC/ATSDR workshop. *MMWR Morb Mortal Wkly Rep*. 1993;42:1–17.
13. *Report of Results of the OSHPD Reabstracting Project*. Sacramento, Calif: Office of Statewide Planning and Development, California Department of Health; 1990.
14. Kaufman GR. *SPARCS Longitudinal File—Concise Summary of Creation Methodology*. Albany, NY: New York State Department of Health; 1989. Memorandum.
15. Blustein J. High-technology cardiac procedures: the impact of service availability on service use in New York State. *JAMA*. 1993;270:344–349.
16. Blustein J. *High-Technology Cardiac Procedures: The Impact of Service Availability on Service Use*. New York, NY: New York University; 1993. Dissertation.
17. 1984 revision of the Uniform Hospital Discharge Data Set. *Federal Register*. July 31, 1985;50:31038–31039.
18. Fleiss JL. *Statistical Methods for Rates and Proportions*. New York, NY: John Wiley and Sons; 1973.
19. Hahn RA, Mulinare J, Teutsch SM. Inconsistencies in coding of race and ethnicity between birth and death in US infants: a new look at infant mortality, 1983 through 1985. *JAMA*. 1992;267:259–263.
20. Massey JT. A comparison of interviewer observed race and respondent reported race in the National Health Interview Survey. *Am Stat Assoc Soc Statistics Section Proc*. 1980;425–428.
21. Kelsey JL, Thompson WD, Evans AS. *Methods in Observational Epidemiology*. New York, NY: Oxford University Press; 1986.
22. Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease relationships and methods of correction. *Annu Rev Public Health*. 1993;14:69–93.
23. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke and coronary artery disease: part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet*. 1990;335:765–774.
24. Mertens TE. Estimating the effects of misclassification. *Lancet*. 1993;342:418–421.
25. Cooper R. Perspective of a health scientist: use of race in public health surveillance. *MMWR Morb Mortal Wkly Rep*. 1993;42:11.

A2623-1

American Journal of Public Health

July 1994, Vol. 84, No. 7 Established 1911

ED KIANKA
HEALTH REPORTER
CONGRESSIONAL INFORMATION SVC
4520 EAST-WEST HIGHWAY
BETHESDA MD 20814
9375

APHA
American Public Health Association

Featuring Children's Health

Editorials and Annotations

1065 Editorial: Public Health Interventions to Reduce Pediatric AIDS *J. J. Goedert and T. R. Coté*
1066 Editorial: The Children's Agenda for the 1990s and Beyond *S. P. Shelov*
1068 Annotation: Voluntary and Involuntary Capture–Recapture Samples—Problems in the Estimation of Hidden and Elusive Populations *R. Neugebauer and J. Wittes*

Public Health Policy Forum

1070 Editorial: Building Partnerships for Healthy Communities—The Role of the Academic Health Center *J. C. Lashof*
1071 The Academic Health Center and the Healthy Community *J. Naughton and J. E. Vana*
1077 Building Bridges between Schools of Public Health and Public Health Practice *A. Handler, L. A. Schieve, P. Ippoliti, A. K. Gordon, and B. J. Turnock*

Articles

VARIOUS TOPICS

1081 The Effect of Ordinances Requiring Smoke-Free Restaurants on Restaurant Sales *S. A. Glantz and L. R. A. Smith*
1086 The Validity of Self-Reported Smoking: A Review and Meta-Analysis *D. L. Patrick, A. Cheadle, D. C. Thompson, P. Diehr, T. Koepsell, and S. Kinne*
1094 Estimating the Number of HIV-Infected Injection Drug Users in Bangkok: A Capture–Recapture Method *T. D. Mastro, D. Kitayaporn, B. G. Weniger, S. Vanichseni, V. Laosunthorn, T. Uneklabh, C. Uneklabh, K. Chooparnya, and K. Limpakarnjanarat*
1100 Self-Rated Health and Survival: A 7-Year Follow-Up Study of Australian Elderly *J. McCallum, B. Shadbolt, and D. Wang*
1106 Work-Related Traumatic Brain Injury in Washington State, 1988 through 1990 *N. J. Heyer and G. M. Franklin*

CHILDREN'S HEALTH

1110 Maternal-Infant HIV Transmission and Circumstances of Delivery *L. Kuhn, Z. A. Stein, P. A. Thomas, T. Singh, and W.-Y. Tsai*
1116 Substance Use and HIV-Related Sexual Behaviors among US High School Students: Are They Related? *R. Lowry, D. Holtzman, B. I. Truman, L. Kann, J. L. Collins, and L. J. Kolbe*
1121 Longitudinal Tracking of Adolescent Smoking, Physical Activity, and Food Choice Behaviors *S. H. Kelder, C. L. Perry, K.-I. Klepp, and L. L. Lytle*
1127 Low Birthweight at Term and the Timing of Fetal Exposure to Maternal Smoking *E. Lieberman, I. Gremy, J. M. Lang, and A. P. Cohen*
1132 Indicators for Monitoring the Growth of Peruvian Infants: Weight and Length Gain vs Attained Weight and Length *E. G. Piwoz, G. L. de Romaña, H. C. de Kanashiro, R. E. Black, and K. H. Brown*

Public Health Briefs

VARIOUS TOPICS

1139 Ethnic Differences in Midwife-Attended US Births *J. D. Parker*
1142 Exposure to Hair-Coloring Products and the Risk of Multiple Myeloma *L. J. Herrinton, N. S. Weiss, T. D. Koepsell, J. R. Daling, J. W. Taylor, J. L. Lyon, G. M. Swanson, and R. S. Greenberg*

(continued on page 1057)

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

July 1994, Vol. 84, No. 7 ISSN: 0090-0036 Established 1911

1144 A Decision-Analytic Approach to Postexposure Rabies Prophylaxis *S. B. Cantor, R. D. Clover, and R. F. Thompson*

CHILDREN'S HEALTH

1148 Mass Media and School Interventions for Cigarette Smoking Prevention: Effects 2 Years after Completion *B. S. Flynn, J. K. Worden, R. H. Secker-Walker, P. L. Pirie, G. J. Badger, J. H. Carpenter, and B. M. Geller*

1151 Counseling Adolescents for Smoking Prevention: A Survey of Primary Care Physicians and Dentists *D. I. Gregorio*

1154 An Increase in *Haemophilus influenzae* Type B Vaccination among Preschool-Aged Children in Inner-City Los Angeles, 1990 through 1992 *D. P. Ewert, S. Westman, B. Ward, N. Bendaña, R. Halbert, C. M. Vadheim, J. I. Ward, and L. Mascola*

1158 Using a Geographic Information System to Understand Child Pedestrian Injury *M. Braddock, G. Lapidus, E. Cromley, R. Cromley, G. Burke, and L. Banco*

1161 The Compliance of Licensed US Child Care Centers with National Health and Safety Performance Standards *D. G. Addiss, J. J. Sacks, M. Kresnow, J. O'Neil, and G. W. Ryan*

1164 Hearing Impairment among 10-Year-Old Children: Metropolitan Atlanta, 1985 through 1987 *C. D. Drews, M. Yeargin-Allsopp, C. C. Murphy, and P. Decouflé*

1167 Socioeconomic Factors and Height of Preschool Children in the Czech Republic *M. Bobák, B. Kříž, D. A. Leon, J. Dáňová, and M. Marmot*

Public Health Then and Now

1171 The Agrarian Myth and Policy Responses to Farm Safety *T. W. Kelsey*

Other Journal Departments

1058 APHA Masthead
1059 Advertisers' Index
1062 What *AJPH* Authors Should Know
1065 Journal Masthead
1178 Letters to the Editor
1186 Book Corner
1189 Annual Meeting Registration Form and Housing Information
1196 Job Opportunities

American Public Health Association

American Journal of Public Health

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 15th St, NW, Washington, DC 20005; (202) 789-5600; TDD (202) 789-5673. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*
Helen Rodriguez-Trias, MD, *Immediate Past President*
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Leonard M. Schuman, MD, *Vice-President (USA)*
John E. F. Hastings, MD, MPH, *Vice-President (Canada)*
Oscar Feo, MD, MSc, *Vice-President (Latin America)*
Jay H. Glasser, PhD, *Treasurer*
Allan G. Rosenfield, MD, *Chair, Executive Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*
Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*
Carol Easley Allan, PhD, RN (1997)
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Patricia P. Evans, MPH (1994)
Eugene Feingold, PhD, JD, *President*
Jay H. Glasser, PhD, *Treasurer*
Alan R. Hinman, MD, MPH (1995)
Margo D. Kerrigan, MPH (1995)
Joanne Lukomnik, MD (1996), *Vice-Chair*
Steven Uranga McKane, DMD, MPH (1995)
Carlos W. Molina, EdD (1997)
Linda Rae Murray, MD, MPH (1996)
Hila Richardson, DrPH, MPH (1996)
Helen Rodriguez-Trias, MD, *Immediate Past President*
Reed V. Tuckson, MD (1994)
Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*
Frances Kissling, *Chair, Action Board*
George E. Pickett, MD, MPH, *Speaker, Program Development Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*

The *American Journal of Public Health* (*AJPH*), published monthly, is the official Journal of the American Public Health Association, Inc, 1015 15th St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005. *AJPH* allotted $40; *The Nation's Health* allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters. Printed in the United States.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on the first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, *AJPH* Reprints, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the Journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The *American Journal of Public Health* is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

Building Bridges between Schools of Public Health and Public Health Practice

Arden Handler, DrPH, Laura A. Schieve, MS, Pamela Ippoliti, MAT, Audrey K. Gordon, PhD, and Bernard J. Turnock, MD, MPH

ABSTRACT

A 1988 Institute of Medicine report, *The Future of Public Health*, characterized the current public health system as fragmented, particularly with regard to relationships between public health agencies and academic institutions. As one response to the report, the Health Resources and Services Administration established the Center for the Development of Public Health Practice at the University of Illinois to advance linkages between schools of public health and public health agencies. Surveys of schools of public health and of state health agencies were conducted in 1992 to collect baseline data on the practice links between the two. Responses reveal that there is a substantial amount of informal collaboration between them. Formalization of collaborative activities between schools and agencies is beginning to occur and is expected to expand owing to increased focus on public health practice at schools of public health. (*Am J Public Health.* 1994;84:1077–1080)

Introduction

The Institute of Medicine's 1988 report, *The Future of Public Health*,[1] cited many inadequacies in the current public health system and called for more formal interaction between public health academe and public health practice as one resolution to the problem. In response to the challenges of this report, the demands posed by the objectives delineated in *Healthy People 2000*,[2] and the recommendations of the Public Health Faculty/Agency Forum,[3] the University of Illinois at Chicago School of Public Health received funding in 1991 from the Health Resources and Services Administration to create the Center for the Development of Public Health Practice. The Center encourages and supports affiliation between schools of public health and public health agencies.

To provide a baseline portrait of collaboration between these two entities, two surveys were conducted in the winter of 1992, one distributed to schools of public health (school survey) and one distributed to state health agencies (agency survey). This report presents the results of these surveys in an effort to quantify the extent of practice activity at schools of public health since the publication of the Public Health Faculty/Agency Forum report.[3]

In developing the school survey, the recommendations of the Public Health Faculty/Agency Forum[3] guided the identification of the following core sentinel public health practice activities:

1. Public health practice steering/advisory committees
2. Multicomponent written agreements between schools of public health and public health agencies (i.e., contractual agreements or memoranda of understanding that are schoolwide and include two or more practice activities)
3. Faculty practice interactions such as faculty appointments at agencies, agency staff appointments at schools, and temporary faculty/agency exchanges
4. The development of clinical tracks at schools of public health
5. Formal practica relationships, including the joint development and evaluation of practica, the training of agency preceptors for their supervisory role, and the provision of funding for the student's participation in practica
6. Technical assistance from schools to agencies and vice versa
7. Joint research

The school survey distinguished between governmental public health agencies and governmental public health–related agencies. The former were defined as agencies that have as their primary purpose health promotion and disease prevention. These include local and state health departments and federal agencies such as the Centers for Disease Control and Prevention. Health-related agencies were defined as those agencies whose primary mission is *not* health promotion and disease prevention, although their functions relate to health. They include agencies such as departments of human services, mental health, and environmental protection. The results presented here focus mainly on governmental public health rather than on public health–related agencies.

The survey of state public health agencies included questions similar to those asked in the school survey. However, since many questions in the agency survey were aimed at obtaining qualitative information, most of the data presented here are from the school survey.

The school survey was sent to all 27 accredited schools of public health, located in 20 states and Puerto Rico. The

The authors are with the Center for the Development of Public Health Practice, School of Public Health, University of Illinois at Chicago.

Requests for reprints should be sent to Arden Handler, DrPH, Program in Community Health Sciences, School of Public Health, University of Illinois at Chicago, 2035 W Taylor, Chicago, IL 60612.

This paper was accepted January 6, 1994.

TABLE 1—Public Health Practice Activities of Schools of Public Health, by Type of Government Agency

School's Practice Activity	Participate with Health Agency (n = 21)		Participate with Health-Related Agency (n = 21)	
	No. Schools	%	No. Schools	%
Practica	21	100	20	95
Joint research	18	86	14	67
Faculty appointments	20	95	17	81
Agency appointments	11	52	4	19
Temporary agency placements	10	48	4	19
Temporary faculty placements	13	62	4	19
Technical assistance from school	21	100	16	76
Technical assistance from agency	18	86	11	52

TABLE 2—Practice Activities Conducted by Schools of Public Health with Government Public Health Agencies, by Agreement Type[a]

School's Practice Activity	Total No. of Schools (n = 21)	Multi-component (n = 5)	Single Component (n = 16)	No Formal Agreement (n = 20)
Practica	21	3	14	19
Joint research	18	3	8	15
Faculty appointments	20	5	7	15
Agency appointments	11	3	4	7
Temporary agency placements	10	4	3	6
Temporary faculty placements	13	4	3	9
Technical assistance from school	21	5	7	19
Technical assistance from agency	18	5	3	15

[a]Categories are not mutually exclusive.

agency survey was distributed to all state public health agencies in the 50 states, the District of Columbia, Puerto Rico, and the Virgin Islands. Telephone contacts were made with each school and agency to verify that they received the survey and to encourage participation. To increase the response rate, additional calls were then made to schools and agencies that did not reply by the deadline. Twenty-one schools, located in 16 states (77% response rate), and 37 health agencies (70% response rate) responded by the summer of 1992. Calls were made to all schools and agencies that returned surveys when responses needed clarification.

A moderate overlap was seen between responding states and schools; for 14 of the 21 responding schools (67%), the corresponding state health agency also responded. Additionally, state agencies were more likely to respond if a school of public health was located within their state, although this difference was not statistically significant (41% vs 23%, $P = .15$). Schools were not matched to their corresponding state health agencies for a comparative analysis, as it was found that many state agencies had linkages with schools located outside their boundaries and schools often engaged in activities with multiple state health agencies.

School Survey Results

All 21 responding schools reported participation in some practice activities with governmental public health and governmental public health–related agencies (Table 1). As expected, schools reported a higher level of interaction with the former agencies than with the latter ones. With both agency types, the most frequently cited activities were practica placements for students, technical assistance from the school for the agency, and faculty appointments for agency staff. What follows is an in-depth examination of specific sentinel public health practice activities taking place at schools of public health.

Public Health Practice Advisory Committees

Sixty-seven percent of the responding schools reported that they use some mechanism for organizing practice activities: 14% use an officially designated Public Health Practice Steering or Advisory Committee exclusively; 29% use another entity (usually a preexisting committee established for another function); and 24% use some combination of these two.

In 50% of the schools that use a formal entity to guide their practice activities (n = 14), at the time of the survey such committees had been in existence for less than 1 year; importantly, the timing of the inception or involvement of these committees in public health practice activities suggests that the Public Health Faculty/Agency Forum's recommendations have had a positive effect on many schools. However, in most schools, meetings of this advisory body occur less often than once a month; in only 29% does it meet monthly.

Formal (Including Multicomponent) Agreements between Schools and Agencies

Nearly all schools (n = 20) reported participation in practice activities with governmental public health agencies *without* formal written agreements, although many schools have several different arrangements with these agencies that include both formal and informal associations. When looking specifically at type of practice activity (Table 2), it is apparent that most interchanges between schools and governmental public health agencies take place without formal written agreements.

Only five schools at the time of the survey were using schoolwide multicomponent agreements. All five include in those agreements faculty appointments for agency staff and technical assistance from the school to the agency and vice versa (Table 2).

Among the 17 schools using either single-component, multicomponent, or both types of formal written agreements to guide their activities with governmental public health agencies, 15 have agreements with state health agencies, 15 have agreements with local health agencies, and 10 have agreements with federal agencies (Table 3). As expected, the most common collaborative activity is practica placements, but the large amount of technical assistance from schools to agencies at all levels indicates that this is also an important area of collaboration.

TABLE 3—Practice Activities Covered by Schools' Formal Written Agreements,[a] by Level of Governmental Public Health Agency

Practice Activity	Federal (n = 10)	State (n = 15)	Local (n = 15)
Practica	9	10	12
Joint research	6	8	5
Faculty appointments	5	7	6
Agency appointments	3	4	2
Temporary agency placements	4	2	2
Temporary faculty placements	4	2	2
Technical assistance from school	5	9	8
Technical assistance from agency	4	5	4

[a]Includes both multi- and single-component agreements (n = 17).

Strategies for Recruiting Public Health Agency Staff

The most common strategy used to promote the appointment of experienced public health practitioners to faculty positions is the adjunct appointment (n = 20). Joint appointments (not specified as clinical or academic) with a governmental public health or health–related agency are reported to be used by almost half the schools (n = 10). Full-time appointments using regular academic titles (n = 7) or clinical titles (n = 9) are each used by fewer than half the schools.

As a possible measure of more extensive commitment or involvement in public health practice activities, we examined the use of different appointment types by whether the school uses broad, multicomponent agreements. All schools that have such agreements also use clinical titles to bring practitioners into academe. Although nearly all schools *without* multicomponent agreements use adjunct appointments, only 25% of these schools use clinical appointments.

Practica

Only 57% of the schools (n = 12) reported that practica are required of all master of public health or equivalent (defined by each school) students; however, this is possibly an undercount as individual department requirements were not considered. The range of students participating in practica at each school is large (13 to 480 annually); approximately one third of these practica experiences involve governmental public health agencies. Encouragingly, 52% of responding schools have established a formal system of practica with governmental public health agencies, 71% indicated that a practica coordinator has been identified within the school, and 48% indicated that a practica coordinator has been identified within governmental public health agencies. Additionally, nearly all schools (n = 19) reported both joint development and joint evaluation of practica.

At the time of the survey, the majority of schools were not providing financial support for practica; however, most schools prepare preceptors for their supervisory role at least some of the time.

State Health Agency Results

Twenty-seven of the 37 responding state health agencies (73%) report participating in practice activities with an accredited school of public health. In agreement with the school survey, practica placements and the receipt of technical assistance from schools were the most frequently cited cooperative activities.

State health agency respondents were also asked if the temporary appointment of faculty from schools of public health to health department positions would be beneficial and, if yes (n = 35: 29 definitively and 6 unsure), to list the divisions where these appointments would be most helpful. The most frequently cited agency divisions for school-to-agency exchanges were environmental health, epidemiology, statistics, maternal and child health, and health planning and policy.

Comment

Together, the data from these surveys render a baseline picture of public health practice activities occurring at schools of public health; these results can be considered representative, given that the response rate to each survey was at least 70%. Whether public health practice activities are conducted informally by individuals or formally through project-specific agreements or through broad, schoolwide, multicomponent agreements may suggest whether these types of activities are supported at the highest levels at both the schools and the agencies. While informal linkages may be fruitful and productive, formal linkages and, in particular, the use of multicomponent agreements may indicate that the practice-academe relationship has been institutionalized. The responses here reveal a substantial amount of collaboration between schools of public health and public health and public health–related agencies; however, much of the collaborative activity reported at the time of these surveys had not been formalized and was occurring between individuals and units rather than between institutions.

Although these results overall suggest a considerable amount of interaction between schools and agencies, it is likely that collaborative activities between these institutions have actually been underestimated. Because there were a significant number of ambiguous answers in individual survey responses, we followed up all school surveys with a telephone call to clarify areas that were identified as being particularly problematic. These callbacks revealed that often the person completing the survey at a given school had difficulty replying for the entire school, given that practice activities were conducted on a departmental basis. Thus, a certain amount of collaboration probably went unrecorded.

With an increase in public health practice activity expected to have taken place at many schools of public health during 1992 to 1993, partly owing to what appears to be a concerted effort by schools to address the Forum's recommendations and partly owing to a public health practice grants initiative of the Center for the Development of Public Health Practice, it is anticipated that there will be a substantial expansion in

formal collaborative activities between schools of public health and public health agencies. Ultimately, it is expected that, as formal interaction between these entities increases and as additional resources are committed to this kind of activity, the training of the public health work force will be more fully integrated with the practice of public health. The goal is a more effective and more practice-based public health work force that will emerge better able to tackle the public health problems that face us all. □

Acknowledgments

This work was sponsored by the US Department of Health and Human Services, Health Resources and Services Administration, through a Public Health Service Special Project Cooperative Agreement (STC-D 1 D38 AH00046-01) with the University of Illinois at Chicago, School of Public Health, Center for Public Health Practice.

This material was previously presented at the 120th Annual Meeting of the American Public Health Association; October 1992; Washington, DC.

We thank the schools of public health and the state health agencies that participated in and completed these surveys for their time and effort. Additionally, their comments and observations were invaluable in providing insight into the development of subsequent surveys and in helping us more clearly focus our efforts toward fostering relationships between public health agencies and academe.

References

1. Institute of Medicine, Committee for the Study of the Future of Public Health/ Division of Health Care Services. *The Future of Public Health.* Washington, DC: National Academy Press; 1988.
2. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services, Public Health Service; 1991. DHHS publication 91–50212.
3. Sorenson AA, Bialek RG. The Public Health Faculty/Agency Forum. Linking Graduate Education and Practice. Final Report. Gainesville, Fla: University Press of Florida; 1993. Sponsored by the Bureau of Health Professions, Health Resources and Services Administration, and the Public Health Practice Program Office, Centers for Disease Control and Prevention.

The Effect of Ordinances Requiring Smoke-Free Restaurants on Restaurant Sales

Stanton A. Glantz, PhD, and Lisa R. A. Smith, BA

ABSTRACT

Objectives. The effect on restaurant revenues of local ordinances requiring smoke-free restaurants is an important consideration for restaurateurs themselves and the cities that depend on sales tax revenues to provide services.

Methods. Data were obtained from the California State Board of Equalization and Colorado State Department of Revenue on taxable restaurant sales from 1986 (1982 for Aspen) through 1993 for all 15 cities where ordinances were in force, as well as for 15 similar control communities without smoke-free ordinances during this period. These data were analyzed using multiple regression, including time and a dummy variable for whether an ordinance was in force. Total restaurant sales were analyzed as a fraction of total retail sales and restaurant sales in smoke-free cities vs the comparison cities similar in population, median income, and other factors.

Results. Ordinances had no significant effect on the fraction of total retail sales that went to restaurants or on the ratio of restaurant sales in communities with ordinances compared with those in the matched control communities.

Conclusions. Smoke-free restaurant ordinances do not adversely affect restaurant sales. (*Am J Public Health.* 1994;84:1081–1085)

Introduction

As the evidence that environmental tobacco smoke endangers nonsmokers[1–5] has accumulated, more and more communities have restricted or eliminated smoking in public places and workplaces. Several communities have enacted legislation that requires smoke-free restaurants, thereby protecting the public and, particularly, restaurant employees[6] from the toxic chemicals in secondhand tobacco smoke. Such legislation, however, is not in the interests of the tobacco industry because creation of smoke-free restaurants is a highly visible statement that tobacco use is no longer socially acceptable.[7] Thus, tobacco companies have sponsored front organizations like the Beverly Hills Restaurant Association, Restaurants for a Sensible Voluntary Policy on Smoking, Californians for Fair Business Policy, and the California Business and Restaurant Alliance to mobilize restaurants against local smoke-free ordinances.[8,9] This strategy achieved its first success in 1987, when the tobacco industry convinced the Beverly Hills City Council to repeal the first 100% smoke-free restaurant ordinance in California on the basis of undocumented claims that business dropped 30% because of the ordinance.[8,9] Because similar predictions for other cities have been published nationally, voiced repeatedly through public testimony, and regularly printed in news reports, we tested the hypothesis that the passage of a smoke-free restaurant ordinance is accompanied by an immediate significant drop in restaurant sales.[10]

This study analyzes sales tax data for the first 15 US cities to enact smoke-free ordinances affecting restaurants. The California cities of Auburn, Bellflower (which repealed its ordinance in March 1992), Beverly Hills (which amended its ordinance 4 months after it went into force), El Cerrito, Lodi, Martinez, Palo Alto, Paradise, Roseville, Ross, Sacramento, and San Luis Obispo, and the Colorado cities of Aspen, Snowmass Village, and Telluride have had such 100% smoke-free restaurant ordinances in force long enough to assess their effects. We also examined sales tax data from 15 comparison cities similar to the smoke-free cities in population, income, smoking prevalence, and other factors.[11–13] An analysis of restaurant sales as a fraction of total retail sales, and of restaurant sales in cities with smoke-free restaurant ordinances compared with those in similar cities that do not have smoke-free ordinances, shows no significant effects on business.

Methods

Data on taxable restaurant sales and total retail sales were obtained from the California State Board of Equalization[14] and Colorado State Department of Revenue[15] from the first quarter of 1986 through the first or second quarter of 1993 (depending on data availability) for the 15 communities that had smoke-free restaurant ordinances in force. Included were cities whose ordinances were in force for at least four quarters during this period, plus Beverly Hills and Bellflower, Calif, whose ordinances were repealed. Data were also obtained for 15 comparison communities where no such smoke-free

The authors are with the Institute for Health Policy Studies, Department of Medicine, at the University of California, San Francisco.

Requests for reprints should be sent to Stanton A. Glantz, PhD, Division of Cardiology, University of California, San Francisco, San Francisco, CA 94143.

This paper was accepted February 4, 1994.

TABLE 1—Profile of Smoke-Free and Comparison Cities

Smoke-Free and Comparison Cities	Population (1989)[a]	Geographical: Inside Urbanized Area	Geographical: Outside Urbanized Area	Geographical: Rural Nonfarm	Type of Smoking Restriction[b]	Median Household Income (1989)[a]	% of Smokers[c]	Date Ordinance in Effect	No. of Months in Effect[d]
Aspen, Colo	5 049		X		100%	37 467	23.5	10/85	95
Vail, Colo	3 659		X		Some	41 211			
Auburn, Calif	10 592		X		100%	37 272	24.1	10/91	21
Oroville, Calif	11 960		X		None	16 614	23.6		
Beverly Hills, Calif	31 971	X			100%	54 348	21.8	4/87–8/87	5
Santa Monica, Calif	86 905	X			Some	35 997			
Bellflower, Calif	61 815	X			100%	32 711	21.8	6/91–3/92	10
Lakewood, Calif	73 000	X			None	44 700			
El Cerrito, Calif	22 869	X			100%	39 538	22.9	11/91	20
San Pablo, Calif	25 158	X			Some	25 479			
Lodi, Calif	51 874	X			100%	30 739	24.1	11/90	32
Merced, Calif	56 216	X			Some	24 727	25.1		
Martinez, Calif	32 038	X			100%	45 964	22.0	3/92	16
Pleasant Hill, Calif	31 585	X			Some	46 885			
Palo Alto, Calif	55 544	X			100%	55 333	19.7	11/91	20
Mountain View, Calif	67 460	X			None	42 431			
Paradise, Calif	25 408		X		100%	22 954	23.6	8/91	23
Red Bluff, Calif	12 363		X		None	19 474			
Roseville, Calif	44 685	X			100%	39 975	24.1	9/91	22
Chico, Calif	40 076	X			Some	19 005	23.6		
Ross, Calif	2 180	X			100%	84 414	21.6	1/90	42
Tiburon, Calif	7 532	X			None	75 864			
Sacramento, Calif	369 365	X			100%	28 183	25.2	5/92	14
Fresno, Calif	354 202	X			Some	24 923	25.1		
San Luis Obispo, Calif	41 958	X			100%	25 982	18.9	8/90	35
Santa Maria, Calif	61 284	X			Some	29 492			
Snowmass, Colo	1 426			X	100%	39 107	23.5	5/89	51
Breckenridge, Colo	1 285			X	Some	33 259			
Telluride, Colo	1 292			X	100%	31 968	23.5	4/88	63
Steamboat Springs, Colo	6 695		X		Some	29 363			

[a]1990 US Census of Population and Housing.[11]
[b]"Some" refers to no more than 60% seating areas for nonsmokers.
[c]Tobacco Use in California (reported by county)[12] for California and Behavioral Risk Factor Surveillance Study for Colorado (statewide) 1991.[13]
[d]Number of months for which data were available for this study.

ordinance was in force or where no more than 60% seating availability for nonsmokers occurred as a part of an existing ordinance (Table 1). Sales data for Aspen and its comparison city were collected from the first quarter of 1982 because Aspen's ordinance was passed in 1985. Data were recorded for "Eating and Drinking Places" and "Total Retail Sales." Published data for restaurant sales and total retail sales in the city of Paradise for the second, third, and fourth quarters of 1990 and in the city of San Luis Obispo for the fourth quarter of 1990 and first quarter of 1991 were corrected as instructed by the Board of Equalization to account for late-reported data (written communications from Robert Rossi, June 15, 1992, and July 20, 1993).

To account for population growth, inflation, and changes in underlying economic conditions, the fraction (F) of total retail sales at restaurants was computed as follows:

$$F = \frac{\textit{Restaurant Sales}}{\textit{Total Retail Sales}}.$$

If an ordinance adversely affected restaurants, this fraction would be expected to drop when the ordinance was in force. Restaurant sales in cities with ordinances were also compared with sales in comparable cities without ordinances as follows:

$$C = \frac{\textit{Restaurant Sales in City with Ordinance}}{\textit{Restaurant Sales in City without Ordinance}}.$$

Again, if an ordinance adversely affected sales this ratio would be expected to drop.

Data were analyzed with linear regression[16]:

$$y = b_0 + b_t t + b_L L + b_w W,$$

where y is the dependent variable (F or C), t is time needed to represent the

underlying secular trend, and L is a dummy variable that indicates whether a smoke-free restaurant law is in force. The estimate of the coefficient b_t quantifies the annual rate of increase (or decrease) in the dependent variable, y, each year. The dummy variable L quantifies the presence of a smoke-free restaurant ordinance as follows:

$$L = \begin{cases} 0 & \textit{If No Ordinance} \\ 1/3 & \textit{If Ordinance in Force for 1 Month of Quarter} \\ 2/3 & \textit{If Ordinance in Force for 2 Months of Quarter} \\ 1 & \textit{If Ordinance in Force for Entire Quarter} \end{cases}$$

The coefficient b_L quantifies the magnitude of the effect of the ordinance on the dependent variable. Because all of the Colorado cities under study are ski centers, the restaurant business is much stronger during the winter tourist season. To allow for this effect, the dummy variable W was included for the Colorado cities, set to 1 for the first quarter (the winter tourist season) and 0 otherwise.

Not only were data analyzed for each city separately, but all the data on restaurant sales as a percentage of total retail sales for all 15 cities with ordinances for the entire year period were pooled in a single analysis, including 29 additional dummy variables, to allow for between-city differences in the mean values of the fraction of total retail sales going to restaurants.

The variance inflation factors for each variable were computed to assess multicollinearity, and the Durbin-Watson statistic was computed to test for autocorrelation among the residuals. The variance inflation factors were always well below 2, and the Durbin-Watson statistic never reached statistical significance. A change is considered statistically significant when $P < .05$.

Results

Table 2 summarizes the results for total restaurant sales as a fraction of all retail sales (F), and total restaurant sales in cities with ordinances compared with those in the matched comparison cities (C). The first column in the table is the mean value observed from 1986 (1982 for Aspen) to the second quarter of 1993 to provide a comparison with the magnitude of the change associated with the ordinance.

Smoke-free ordinances generally had no statistically significant effect on the fraction of retail sales that went to restaurants or on total restaurant sales in cities with ordinances compared with those in cities without smoke-free ordinances (Table 2 and Figure 1). There is marginal evidence that the fraction of total retail sales to restaurants increased in two cities (Bellflower, $P = .025$; Martinez, $P = .008$) and decreased in one city (Roseville, $P = .039$). In a comparison of restaurant sales in one city with an ordinance versus one city without an ordinance, sales increased in one city (Palo Alto, $P = .004$) and decreased in another (Paradise, $P = .049$). The lack of consistent response suggests that these results may simply reflect random variation, given the large number of P values that were computed. Analysis of all the data in pooled regressions did not detect significant changes in the percentage of retail sales or sales in cities with smoke-free ordinances compared with those in cities without ordinances.

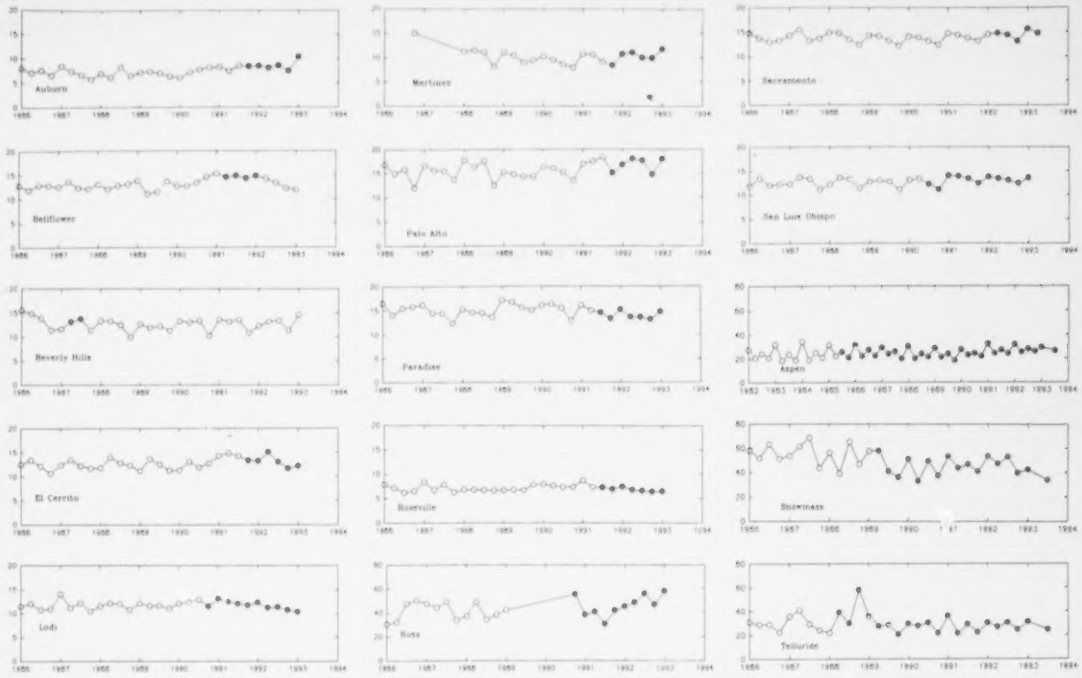
TABLE 2—Effect of Smoke-Free Restaurant Ordinances on Total Restaurant Sales

City	Mean	Effect of Ordinance: Change, b_L	Effect of Ordinance: P	Model: R^2	Model: P
Fraction of total retail sales, F, %					
Aspen	24.8	1.1 ± 1.3	.408	.688	.000
Auburn	7.5	1.0 ± 0.5	.092	.319	.007
Bellflower	13.1	1.5 ± 0.6	.025	.313	.008
Beverly Hills	12.8	0.6 ± 1.2	.633	.033	.646
El Cerrito	12.7	−0.4 ± 0.7	.637	.100	.255
Lodi	11.7	0.1 ± 0.6	.902	.005	.939
Martinez	10.3	2.9 ± 1.0	.008	.404	.007
Palo Alto	15.8	0.7 ± 1.1	.520	.115	.204
Paradise	14.9	−1.4 ± 0.8	.078	.181	.075
Roseville	7.1	−0.9 ± 0.4	.039	.156	.111
Ross	43.5	−3.3 ± 9.1	.715	.132	.243
Sacramento	13.9	0.9 ± 0.6	.102	.102	.233
San Luis Obispo	12.7	0.2 ± 0.6	.764	.082	.327
Snowmass	49.2	6.0 ± 5.7	.301	.374	.006
Telluride	29.6	9.4 ± 4.7	.055	.197	.120
All combined	18.4	−1.3 ± 1.0	.210	.611	.000
Ratio of sales with comparison city, C					
Aspen	1.12	.21 ± .12	.106	.153	.071
Auburn	.44	.03 ± .02	.186	.082	.327
Bellflower	.50	−.02 ± .02	.347	.036	.621
Beverly Hills	.56	−.06 ± .04	.171	.238	.029
El Cerrito	1.28	−.00 ± .08	.998	.053	.495
Lodi	.90	−.01 ± .03	.742	.270	.017
Martinez	.41	.04 ± .03	.194	.329	.001
Palo Alto	1.69	.23 ± .07	.004	.416	.001
Paradise	.71	−.07 ± .03	.049	.144	.132
Roseville	.68	−.02 ± .03	.562	.089	.300
Ross	.05	.02 ± .01	.196	.302	.028
Sacramento	1.10	−.05 ± .03	.091	.403	.001
San Luis Obispo	1.12	−.08 ± 0.6	.177	.154	.113
Snowmass	.95	−.29 ± .20	.193	.584	.000
Telluride	.42	.08 ± .07	.282	.372	.006
All combined	.82	−.04 ± .03	.166	.828	.000

Beverly Hills is a particularly important case because it has been used by the tobacco industry to support the claim that smoke-free restaurant ordinances are associated with a 30% drop in business (Figure 2). However, data reveal that no such drop in sales occurred upon enactment, and that no increase in sales followed repeal 4 months later. Likewise, despite the fact that the Bellflower ordinance was repealed because of claims that business dropped, the ordinance was actually associated with a marginally significant ($P = .025$) increase in business.

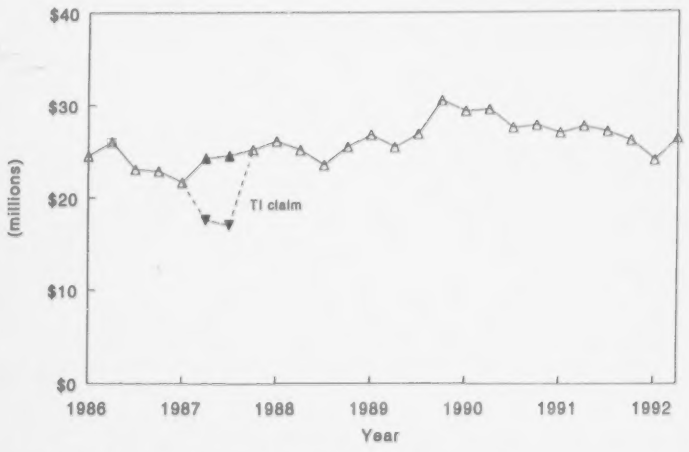
Discussion

This is the first comprehensive study that examines taxable sales data to determine the economic impact of smoke-free restaurant ordinances on restaurant sales. Using data from the California State Board of Equalization and the Colorado State Department of Revenue for pur-



Note. The quarters in which 100% smoke-free ordinances were in effect are represented by solid circles.

FIGURE 1—Restaurant sales as a percentage of total retail sales for the 15 communities included in the study.



Note. Period of smoke-free ordinance is indicated by the solid triangles.

FIGURE 2—The 100% smoke-free restaurant ordinance in force in Beverly Hills did not reduce sales by 30% (dashed line with "TI [tobacco industry] claim"), as the tobacco industry had suggested; rather, it had no significant effect on sales.

poses of paying sales taxes has several advantages. First, the numbers reflect *all* restaurant sales in a community, not just those of a small sample of restaurants. Second, the numbers are objective; they were collected through consistent methods by agencies with no interest in the effects of smoking restrictions on restaurant sales. Third, sales tax data can be expected to be reasonably accurate since it is a crime to lie in reporting the figures.

The communities studied in the report are different from each other and represent a cross-section of communities that might enact legislation controlling smoking in restaurants: Auburn is a small Sierra foothills community; Beverly Hills is a well-to-do urban city; Bellflower is a middle-class bedroom community; El Cerrito and Martinez lie within highly industrial areas; Lodi is a rural agricultural center; Palo Alto is a large suburban university community; Paradise is a small, semiagricultural community; Sacramento is a large city and the state capital; San Luis Obispo is a college town; Roseville represents a semirural bedroom commu-

nity; Ross is a small affluent San Francisco Bay community; and the three Colorado cities are mountainous, tourist resort areas. The fact that there were no adverse effects on business in these communities supports the conclusion that the results generalize broadly. Further, these 15 cities represent every city that has passed smoke-free ordinances that have been in effect long enough to study.

This study covers a significant period of time. It is important to take into account long-term (secular) trends as well as the quarter-by-quarter random variation and short-term economic changes. We avoided short-term analyses because it is generally possible to reach any conclusion desired by selectively picking the "correct" two quarters for analysis.[17] To avoid such biases and increase the power of the statistical analysis to detect an effect of the ordinances, we used data for a 7-year period (12 years for Aspen and Vail). This length of time allowed us to obtain good estimates of secular trends before evaluating any effects of the ordinances.

A common concern is raised about the possibility that patrons will dine in adjacent communities without such restrictions. Our data address this concern because the cities examined in this study are not isolated communities. Auburn, Lodi, Martinez, Paradise, Roseville, and San Luis Obispo, while not in large urban centers, are all surrounded by unincorporated areas that contain restaurants. Beverly Hills and Bellflower and their comparison cities, Santa Monica and Lakewood, are all located in Los Angeles County, a major metropolitan area in which all communities directly abut other communities. El Cerrito, Palo Alto, Ross, and their comparison cities all lie within the San Francisco Bay region. Sacramento and its comparison city, Fresno, both large urban centers, face competition from several neighboring communities. Although the skiing communities of Aspen, Telluride, and Snowmass Village are relatively secluded, other resort towns nearby that allow smoking would represent viable tourist alternatives to these smoke-free cities. If people were leaving these cities to dine in neighboring cities, our analysis would have detected it.

Another area of concern is the effect on bars since smoking and drinking are thought to go together. Revenues from bars and "full-service" restaurants are included in the sales tax data we used. The ordinances examined in this study contain different provisions governing bars independently and bars in relation to restaurants. Had there been a significant effect on sales in such restaurants, our analysis would have detected it. Furthermore, an analysis of individual classes of restaurants (based on whether they sell different types of alcohol) for four cities in California previously showed no effect when full-service restaurants were analyzed separately.[18]

Finally, the fact that the ordinances in Beverly Hills and Bellflower were repealed adds to the strength of our conclusions. Had the ordinances affected sales negatively, we would have expected to see an increase in sales following repeal. However, there was no increase in Beverly Hills, and sales dropped in Bellflower after the ordinance was repealed. Thus, legislators and government officials can enact such health and safety requirements to protect patrons and employees in restaurants from the toxins in secondhand tobacco smoke without the fear of adverse economic consequences. □

Acknowledgments

This research was supported by funds provided by the Cigarette and Tobacco Surtax Fund of the State of California through the Tobacco-Related Disease Research Program of the University of California (award 1RT 520).

References

1. Centers for Disease Control, Office on Smoking and Health. *The Health Consequences of Involuntary Smoking: A Report of the Surgeon General.* Rockville, Md: US Dept of Health and Human Services, Public Health Service; 1986. DHHS publication CDC 87–8398.
2. National Research Council, Committee on Passive Smoking. *Environmental Tobacco Smoke: Measuring Exposures and Assessing Health Effects.* Washington, DC: National Academy Press; 1986.
3. *Respiratory Health Effects of Passive Smoking: Lung Cancer and Other Disorders.* Washington, DC: Environmental Protection Agency, Office of Health and Environmental Assessment; 1992. Publication EPA/600/6–90/006F.
4. Glantz SA, Parmley WW. Passive smoking and heart disease: epidemiology, physiology, and biochemistry. *Circulation.* 1991;83: 1–12.
5. Steenland K. Passive smoking and the risk of heart disease. *JAMA.* 1992;267:94–99
6. Siegel M. Involuntary smoking in the restaurant workplace: a review of employee exposure and health effects. *JAMA.* 1993;270:490–493.
7. *Public Attitudes toward Cigarette Smoking and the Tobacco Industry.* New York, NY: Roper Organization; May 1978.
8. Samuels B, Glantz S. The politics of local tobacco control. *JAMA.* 1991;266:2110–2117.
9. Traynor MP, Begay ME, Glantz SA. The new tobacco industry strategy to prevent local tobacco control. *JAMA.* 1993;270:479–486.
10. Press Alert and Fact Sheet: San Francisco's Hospitality Industry Faces Smoking Ban. San Francisco, Calif: Golden Gate Restaurant Association; February 17, 1992.
11. *United States Census of Population and Housing 1990: Summary Tape File 3A—Bureau of the Census.* Washington, DC: US Bureau of the Census; 1990.
12. Pierce JP, Farkas A, Evans N, et al. *Tobacco Use in California 1992. A Focus on Preventing Uptake in Adolescents.* Sacramento, Calif: California Department of Health Services; 1993.
13. *Behavioral Risk Factor Surveillance System—Summary Prevalence Report.* Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion; 1991.
14. *Taxable Sales in California* (*Sales and Use Tax*). Sacramento, Calif: State of California, Board of Equalization; Table 5 (quarterly reports 1986–1992).
15. *Sales Tax Statistics for Counties and Selected Cities and Detailed Industry Reports by County (January 1982–June 1992).* Denver, Colo: State of Colorado, Department of Revenue.
16. Mendenhall W, Sincich T. *A Second Course in Business Statistics: Regression Analysis.* San Francisco, Calif: Dellen Publishing Company; 1989: chap 9.
17. *Impacts of Local 100% Smoking Bans on Restaurants and Related Businesses.* Los Angeles, Calif: California Business and Restaurant Alliance; 1992.
18. Glantz SA, Smith LRA. *The Effect of Ordinances Requiring Smoke-free Restaurants on Restaurant Sales in California.* San Francisco, Calif: University of California Institute for Health Policy Studies Monograph Series; 1992.

The Validity of Self-Reported Smoking: A Review and Meta-Analysis

ABSTRACT

Objectives. The purpose of this study was to identify circumstances in which biochemical assessments of smoking produce systematically higher or lower estimates of smoking than self-reports. A secondary aim was to evaluate different statistical approaches to analyzing variation in validity estimates.

Methods. Literature searches and personal inquiries identified 26 published reports containing 51 comparisons between self-reported behavior and biochemical measures. The sensitivity and specificity of self-reports of smoking were calculated for each study as measures of accuracy.

Results. Sensitivity ranged from 6% to 100% (mean = 87.5%), and specificity ranged from 33% to 100% (mean = 89.2%). Interviewer-administered questionnaires, observational studies, reports by adults, and biochemical validation with cotinine–plasma were associated with higher estimates of sensitivity and specificity.

Conclusions. Self-reports of smoking are accurate in most studies. To improve accuracy, biochemical assessment, preferably with cotinine–plasma, should be considered in intervention studies and student populations. (*Am J Public Health.* 1994;84:1086–1093)

Donald L. Patrick, PhD, MSPH, Allen Cheadle, PhD, Diane C. Thompson, MS, Paula Diehr, PhD, Thomas Koepsell, MD, MPH, and Susan Kinne, PhD

Introduction

Smoking continues to be the largest single preventable cause of premature mortality and morbidity in the United States, yet 29% of American adults continue to smoke.[1] Efforts to promote cessation of smoking include interventions conducted with patients in clinical practices, in group environments such as schools and work sites, and in entire communities. Self-reports of smoking behavior are often assessed to determine the efficacy of these interventions. Observational studies and epidemiological studies of risk also incorporate measures of smoking behavior. Smoking is assessed in these studies to discriminate between smokers and nonsmokers, to measure change in smoking status, or to calculate pack-years of exposure retrospectively for risk assessment.

The validity of self-reported smoking is often questioned because of the widespread belief that smokers are inclined to underestimate the amount smoked[2,3] or to deny smoking at all.[4,5] As more attention is paid to smoking in the media and in public places, work sites, and clinical practice, individuals become sensitized to socially desirable forms of behavior. Thus, smokers may be more likely to exaggerate the extent to which their behavior conforms to the perceived social norm of "not smoking." Bias may be more common wherever social desirability is greater, such as in community-based studies in which intervention programs often seek explicitly to change community norms about the social acceptability of smoking.

Biochemical assessments of smoking by-products in body substances are often made to validate self-reports of smoking. Biochemical assessments can be viewed primarily as measures of the point prevalence of current smoking.[6] Because they are believed to be more objective and less susceptible to bias, biochemical measures are most often considered the "gold standard" in validation studies (i.e., they are considered more accurate than self-reports of smoking). Cotinine (in plasma, saliva, or urine), thiocyanate (in plasma or saliva), and carbon monoxide (in expired air) are the most commonly used biochemical assessments. Participants are either told in advance that such assessments will be made or asked for informed consent and specimens "on the spot." Sometimes the bogus pipeline procedure has been used, wherein subjects are informed that their self-reports can or will be objectively verified by the researchers by means of a biochemical test. In actuality, no verification takes place, although specimens are collected and left unanalyzed.[4]

Despite their believed objectivity, biochemical measures do not provide a gold standard, nor are they perfect measures of accuracy for use in assessing criterion validity. Carbon monoxide and thiocyanate can be elevated in those who do not use tobacco, and cotinine, although a specific metabolite of nicotine, can be elevated in users of snuff and chewing tobacco. When biochemical tests are

Donald L. Patrick, Allen Cheadle, Paula Diehr, and Thomas Koepsell are with the Department of Health Services, University of Washington, Seattle. Donald L. Patrick, Thomas Koepsell, Diane C. Thompson, and Susan Kinne are with the Fred Hutchinson Cancer Center, Seattle. Paula Diehr is also with the Department of Biostatistics, and Thomas Koepsell with the Department of Epidemiology, University of Washington.

Requests for reprints should be sent to Donald L. Patrick, PhD, MSPH, Department of Health Services, SC-37, University of Washington, Seattle, WA 98195.

This paper was accepted October 27, 1993.

repeated, the results may be different even when smoking status has not changed. Biochemical measures also have practical drawbacks. Although nonreactive, these measures are obtrusive: blood, saliva, or breath samples need to be collected from the individual. Collection of samples involves more contact with respondents than usual in conducting large-scale field studies[3] and may result in increased refusals.[6] Because of the short half-life of smoking by-products in the body, biochemical assessment validates only smoking status near the time of specimen collection.[7] Costs can be considerable, ranging from less than $1 per sample for carbon monoxide to $20 per sample for cotinine analysis.[8] The cost of collecting, handling, and arranging for frozen storage of the specimens can add significantly to these estimates.

In contrast, self-reported smoking is assessed easily by using self-administered questionnaires in person or by mail or by using interviewer-administered questionnaires in person or on the phone. Questionnaires are noninvasive and inexpensive, and assurances of confidentiality of information can reduce refusals to participate. Self-reported information can be used to measure behavioral change, to calculate exposure risk, or to study pathways to smoking cessation or continuation.

The meta-analyses reported in this paper combine findings from a number of studies that validated self-reported smoking with biochemical measures, making it possible to examine the importance of different aspects of the studies, the populations, and the validation process. This paper addresses four major questions. First, what evidence exists to document the validity of self-reported measures of smoking behavior? Second, under what circumstances is it most important for investigators to consider biochemical assessment in studies of smoking behavior? Third, how do results change when using different statistical approaches for analyzing variation in the measures of accuracy and for pooling information across studies? Finally, how does this literature review inform the conduct of future validation studies, reports of smoking behavior, and the publication of results?

Methods

Description of Meta-Analysis Procedures

Meta-analyses are becoming common in both clinical and social science research.[9–11] Meta-analytic techniques were applied in this study to observations of the association between biochemical measures and self-reported smoking, similar to meta-analyses of diagnostic tests. Similar applications, such as the accuracy of the exercise electrocardiogram[12] and the human immunodeficiency virus antibody test,[13] have appeared in the literature. This application, like meta-analyses of randomized clinical trials, is an observational study of previously published studies.

Standard procedures were followed in accumulating and evaluating research studies for the meta-analysis.[9–11] We defined the problem as accuracy of self-reported smoking, with biochemical assessment as the criterion or concordance measure for evaluating validity. Bibliographic searches were conducted on all articles published between 1982 and 1991. Initially the MEDLINE database was used, with "smoking" as the subject head and the keywords "intervention studies," "evaluation," "community-based programs," and "education" as subheads. The Science Citation Index was used to trace articles referenced in studies previously identified in the bibliographic search. The Current Contents database was scanned for more recent articles through mid-1991, and references in these articles were also evaluated. Investigators familiar with smoking research at the Fred Hutchinson Cancer Research Center were also asked to identify appropriate studies from their literature files.

Thirty studies containing comparisons between self-reported smoking and biochemical assessments were identified. Studies confined to pregnant women were excluded. All studies were reviewed for information on the following characteristics: method of administration (self-administered vs interviewer administered), biochemical measures (cotinine, thiocyanate, or carbon monoxide), type of sample (air, blood, saliva), cutoff value used for biochemical definition of smoking, population (student vs general population), study design (intervention vs observational), sample size, ability to classify participants according to a 2 × 2 table based on self-report of smoking (yes/no) and the gold-standard definition of exceeding the defined cutoff level on the biochemical measure, and the smoking rate (i.e., prevalence of smoking), defined by the gold standard and self-report measures. The 2 × 2 table for calculating the accuracy of reports is shown in Figure 1.

		Respondent Exceeds Cut-Off on Biochemical Measure?	
		Yes	No
Self-Reported Smoking?	Yes	a	b
	No	c	d

FIGURE 1—The 2 × 2 table.

All the required data were available in 26 of the 30 articles identified as validity studies of self-reported smoking.[4,5,14–37] Four studies did not contain sufficient information to calculate accuracy measures, and these studies were eliminated from further analysis. Three members of the study team (Donald L. Patrick, Diane C. Thompson, and Susan Kinne) abstracted data independently to ensure quality control of the data used in the analyses. Discrepancies among the three abstractors were investigated and resolved after discussion. The 26 studies contained 32 comparisons based on independent samples and 51 comparisons wherein 2 or more comparisons were made on partial or total analyses of the same individuals. Table 1 contains the essential data abstracted from the studies included in the meta-analysis.

Measures of Accuracy

Data abstracted from the 26 studies were used to calculate two measures of self-report accuracy. For the purposes of this study, in which biochemical measures were considered the criterion measure, *sensitivity* was defined as $a/(a + c)$ in Figure 1, or the proportion of respondents with a positive level on the biochemical measure that reported smoking. *Specificity* was defined as $d/(b + d)$, or the proportion of respondents with a negative level on the biochemical measure that reported absence of smoking.

Analytic Models and Procedures

Cutoff levels were standardized into comparable units of measurement across studies for each type of biochemical measure and biological specimen. Studies using thiocyanate–saliva samples were eliminated from our analyses because of their outlying values for sensitivity and specificity after such standardization. We thus analyzed 47 of the 51 comparisons (allowing more than 1 comparison per

TABLE 1—Data Used in Meta-Analysis of 26 Published Studies on Biochemical Validation of Self-Reported Smoking

Author(s) (Publication Year)	Study No.	Comparison No.	Method of Administration	Biochemical Measure	Sample	Cutoff Value	Population	Study Design	Sample Size	a + c[a]	Sensitivity	b + d[b]	Specificity	Gold Standard Smoking Rate	Self-Report Smoking Rate
Bauman & Dent (1982)[14]	1	1	SAQ	CO	Air	9	S	O	389	36	58	352	92	9.3	12.6
Bauman & Dent (1982)	1	2	SAQ	CO	Air	9	S	O	1189	100	90	1089	89	8.4	17.7
Bauman & Dent (1982)	1	3	SAQ	CO	Air	9	G	O	370	112	93	258	90	30.3	35.1
Bauman & Dent (1982)	1	4	SAQ	CO	Air	9	G	O	1115	350	95	765	88	31.4	38.1
Bauman et al. (1982)[15]	2	1	SAQ	CO	Air	9	S	O	82	3	100	79	97	3.7	6.1
Bauman & Koch (1983)[16]	3	1	SAQ	CO	Air	9	G	O	1439	460	95	979	92	32.0	36.3
Bauman & Koch (1983)	3	2	SAQ	CO	Air	9	G	O	386	36	64	350	95	9.3	10.9
Cohen & Bartsch (1980)[17]	4	1	SAQ	COHb	Blood	2	G	O	617	385	91	232	67	62.4	69.0
Cohen & Bartsch (1980)	4	2	SAQ	SCN	Blood	100	G	O	617	431	92	186	83	69.9	69.0
Fortman et al. (1984)[18]	5	1	IAQ	CO	Air	9	G	O	1279	416	91	863	99	32.5	30.4
Fortman et al. (1984)	5	2	IAQ	SCN	Blood	100	G	O	1673	550	87	1123	96	32.9	31.4
Gillies et al. (1982)[19]	6	1	SAQ	SCN	Saliva	100	S	O	252	48	6	204	97	18.7	3.6
Haddow et al. (1988)[20]	7	1	SAQ	COT	Blood	10	G	O	296	64	92	232	98	21.3	21.3
Luepker et al. (1981)[21]	8	1	SAQ	SCN	Saliva	100	S	O	1205	74	37	1258	98	5.0	3.5
Luepker et al. (1989)[5]	9	1	SAQ	COT	Saliva	20	S	O	263	88	92	175	97	33.5	33.1
Luepker et al. (1989)	9	2	SAQ	COT	Saliva	20	S	O	290	108	95	182	93	37.2	39.7
Luepker et al. (1989)	9	3	SAQ	COT	Saliva	20	S	O	325	105	77	220	97	32.3	26.8
Luepker et al. (1989)	9	4	SAQ	COT	Saliva	20	S	O	331	111	93	220	95	33.5	34.7
McNeill et al. (1987)[22]	10	1	SAQ	COT	Saliva	15	S	O	508	125	96	383	86	24.6	34.1
Murray et al. (1987)[24]	11	1	SAQ	CO	Air	9	S	O	245	46	46	199	89	18.8	17.6
Murray et al. (1987)	11	2	SAQ	CO	Air	9	S	O	248	15	100	233	85	6.0	20.6
Murray et al. (1987)	11	3	SAQ	CO	Air	9	S	O	246	19	89	227	88	7.7	18.3
Noland et al. (1988)[23]	12	1	SAQ	COT	Saliva	25	S	O	308	121	99	187	79	39.0	51.6
Pechacek et al. (1984)[24]	13	1	SAQ	SCN	Saliva	100	S	O	1183	184	72	999	89	15.6	20.8
Pechacek et al. (1984)	13	2	SAQ	CO	Air	9	S	O	1221	166	95	1055	91	13.6	20.7
Petitti et al. (1981)[25]	14	1	SAQ	SCN	Blood	100	G	O	267	55	100	212	85	20.6	32.2
Petitti et al. (1981)	14	2	SAQ	CO	Air	9	G	O	267	58	98	209	86	21.7	32.2
Pierce et al. (1987)[26]	15	1	IAQ	COT	Saliva	40.5	G	O	975	336	93	639	93	34.5	36.2
Pojer et al. (1984)[27]	16	1	SAQ	COHb	Blood	2	G	O	368	166	98	202	88	45.1	50.8
Pojer et al. (1984)	16	2	SAQ	COT	Blood	40.5	G	O	368	181	98	187	95	49.2	50.8
Pojer et al. (1984)	16	3	SAQ	SCN	Blood	70	G	O	369	144	98	225	80	39.0	50.7
Prignot (1987)[28]	17	1	IAQ	SCN	Blood	80	G	O	197	31	90	166	99	15.7	15.2
Saloojee et al. (1982)[29]	18	1	IAQ	COHb	Blood	1.6	G	O	439	349	99	90	86	79.5	82.0
Saloojee et al. (1982)	18	2	IAQ	SCN	Blood	73	G	O	439	339	99	100	75	77.2	82.0
Sillett et al. (1978)[30]	19	1	SAQ	COHb	Blood	1.7	G	I	79	36	69	43	93	45.6	35.4
Sillett et al. (1978)	19	2	SAQ	COHb	Blood	1.7	G	I	140	89	63	51	96	63.6	41.4
Slattery et al. (1989)[31]	20	1	IAQ	COT	Blood	15	G	O	542	157	95	385	96	29.0	30.1
Stookey et al. (1987)[32]	21	1	IAQ	COT	Saliva	10	G	I	338	260	82	78	97	76.9	63.9
Stookey et al. (1987)	21	2	IAQ	CO	Air	9	G	I	338	234	88	104	91	69.2	63.9
Stookey et al. (1987)	21	3	IAQ	COT	Saliva	10	G	I	236	214	100	22	91	100.0	99.0
Stookey et al. (1987)	21	4	IAQ	SCN	Saliva	80	G	I	236	193	97	43	33	96.8	86.5
Stookey et al. (1987)	21	5	IAQ	CO	Air	9	G	I	236	207	100	29	69	100.0	95.8
Van Vunakis et al. (1989)[33]	22	1	IAQ	COT	Blood	8	G	O	318	137	92	181	100	43.1	39.6
Van Vunakis et al. (1989)	22	2	IAQ	COT	Saliva	25	G	O	303	122	94	181	99	40.3	38.6
Vogt et al. (1977)[34]	23	1	IAQ	CO	Air	9	G	I	123	79	97	44	98	64.2	63.4
Vogt et al. (1977)	23	2	IAQ	CO	Air	9	G	I	139	85	98	54	80	61.2	67.6
Vogt et al. (1977)	23	3	IAQ	SCN	Blood	100	G	I	123	74	96	49	86	60.2	63.4
Vogt et al. (1977)	23	4	IAQ	SCN	Blood	100	G	I	139	79	96	60	70	56.8	67.6
Wagenknecht et al. (1990)[35]	24	1	IAQ	COT	Blood	13	G	O	4932	1542	88	3390	98	31.3	28.9
Wald et al. (1981)[36]	25	1	IAQ	COHb	Blood	2	G	O	8724	1666	99	7055	94	19.1	23.9
Williams et al. (1979)[37]	26	1	SAQ	COT	Blood	3	S	O	118	21	90	97	99	17.8	16.9

Note. SAQ = self-administered questionnaire; IAQ = interviewer-administered questionnaire. CO = carbon monoxide measured in parts per million; COHb = carboxyhemoglobin, measured in percent; SCN = thiocyanate measured in micromoles per liter; COT = cotinine measured in nanograms per milliliter. S = student; G = general population; O = observational; I = intervention.

[a]Respondent exceeded cutoff on biochemical measure (see Figure 1).

[b]Respondent did not exceed cutoff on biochemical measure (see Figure 1).

study) and 30 of the 32 strictly independent comparisons (allowing only 1 comparison per study, chosen at random).

Since there is uncertainty about how best to combine proportions across studies,[40] several analytic methods were used to obtain point estimates and confidence intervals of average sensitivity and specificity. The fixed-effects approach assumes that all studies are homogeneous and that the "true" sensitivity and specificity are the same for each study. Under this model, estimates from each study differ from this assumed true value only because of sampling variation, and the estimates from small studies tend to be worse than those from larger studies. In the fixed-effects model, the most efficient point estimate for sensitivity, for example, is the mean of the sensitivities weighted by sample size. This is equivalent to putting all of the people from all studies into one 2×2 table and estimating sensitivity and specificity from it. Confidence intervals are calculated from the weighted standard error. Regression analyses to determine correlates of accuracy can be achieved by using weighted ordinary least squares, with the study as the unit of analysis, or by using logistic regression. We used logistic regression to conduct fixed-effects analyses of the correlates of accuracy.

In actuality, however, studies are not homogeneous: they involve different study populations and different data collection methods. The "true" sensitivity and specificity values being estimated probably do vary among studies because of this heterogeneity. Under a random-effects model, variation among the studies is assumed to be due, in part, to the variation among the "true" values, as well as to sampling variation within each study. The correct weight in this case depends on both sources of variation. If the between-studies heterogeneity is large, this results in approximately equal weights for each study; if it is small, then the weights are approximately proportional to the sample size, as in the fixed-effects model. Laird and Mosteller[38] suggested a moment-based estimate for the mean, with its associated variance, that can be used to obtain confidence intervals for the random-effects method. This method computes appropriate weights for each study; these weights are a combination of the variance within studies and the variances among studies. If there is almost no variation among studies, then each study is weighted by its sample size. If there is great variance among studies, then each study will receive approximately equal weight. We computed the weights and found that the latter case was true for our data. The sensitivity and specificity estimates for our studies were so variable that the random-effects analysis essentially gave all studies equal weight. For this reason, to examine the effects of covariates under a random-effects assumption, we used simple ordinary least squares regression with the study as the (unweighted) unit of analysis and sensitivity and specificity as the dependent variables.

The sensitivity and specificity measures were negatively skewed. To test the sensitivity of the findings to this deviation from normality, we analyzed the logarithm of 100 minus sensitivity plus 1 and 100 minus specificity plus 1, which did have a reasonably normal distribution. The results of analyses using the logarithmically transformed measures were substantially the same as analyses with untransformed data; hence, for ease of interpretation, we report the original analyses.

We compared fixed-effects and random-effects models. For the purposes of this paper, we present unadjusted, bivariate results and multivariate analyses using the ordinary least squares random-effects regression (most conservative) and the fixed-effects logistic regression (least conservative). Independent variables in the regression analyses were method of administration (self-administered vs interviewer administered), study design (observational vs intervention), population type (student vs general population), and type of biochemical measure and specimen (cotinine–plasma as the reference vs cotinine–saliva, thiocyanate–plasma, or carbon monoxide [air and blood]). Two-way interactions among the study characteristics were examined in the random-effects ordinary least squares analyses of 47 comparisons; one interaction was entered at a time.

Results

A total of 36 830 respondents were included in the 26 studies and 51 comparisons. Of the comparisons reported in these studies, 37.7% were obtained from self-administered questionnaires and 62.3% were obtained by interviewers. Students represented 22.8% of respondents for the comparisons; 77.2% of respondents were drawn from the general population. Only 5.8% were respondents in intervention studies in which biochemical assessments were used. Carbon monoxide, thiocyanate, and cotinine were used as the biochemical measure in 54.8%, 18.2%, and 27% of the comparisons, respectively.

As shown in Table 1, sensitivity values ranged from 6 to 100 in the 51 comparisons; specificities ranged from 33 to 100. The average sensitivity for these studies, unweighted by sample size, was 87.5; the average specificity was 89.2. Tables 2 and 3 show results for sensitivity and specificity for all 47 comparisons and for the 30 strictly independent comparisons, respectively. For each sample, results are shown for the ordinary least squares random-effects and logistic fixed-effects models. The bivariate analyses show the effect on sensitivity or specificity of each study characteristic by itself. The multivariate analysis results come from regression models that included as predictors all of the study characteristics listed. Thus, they show the effect of each predictor on sensitivity or specificity after adjusting statistically for all of the other predictors.

For the fixed-effects analysis, the logistic regression coefficients were used to calculate the effect, in percentage points, of the study characteristic of interest when other characteristics (if any) in the model were held constant at their mean values. Effects for the ordinary least squares random-effects regression were the coefficients themselves. For example, per the random-effects ordinary least squares regression treating comparisons as independent (Table 2), interviewer-administered survey self-reports had a sensitivity that was 5.2 percentage points higher than that of self-administered survey reports, but this difference was not significantly different from zero ($P = .135$).

For sensitivity, the fixed-effects result for interviewer-administered studies was a 1.1 percentage point increase, which was significant at the .05 level. After controlling for all other study characteristics, the effect was a 4.0 percentage point increase in self-report sensitivity for the random-effects model (not significant) and a 1.2 percentage point increase in the fixed-effects analysis (marginally significant at $P = .066$). None of the study characteristics showed significantly different estimates for sensitivity in either the bivariate or multivariate random-effects analyses. For the fixed-effects approach, student populations and the type of sample were significantly related to sensitivity. Student populations yielded significantly lower sensitivity than general population studies. All biochemical samples yielded higher

TABLE 2—Estimated Effects of Study Characteristics on Sensitivity and Specificity of Self-Reported Smoking: All Comparisons (n = 47)

	Sensitivity				Specificity			
	Ordinary Least Squares Random Effects		Logistic: Fixed Effects		Ordinary Least Squares Random Effects		Logistic: Fixed Effects	
Study Characteristic	Effect[a]	*P*	Effect[a]	*P*	Effect[a]	*P*	Effect[a]	*P*
Bivariate analyses								
Interviewer administered	5.2	.135	1.1	.047	2.5	.313	6.3	<.001
Student population	−4.6	.227	−3.7	<.001	1.8	.510	−3.4	<.001
Observational study	−0.3	.937	−1.0	.135	4.2	.124	4.8	<.001
Cotinine–saliva[b]	−0.4	.944	2.3	.017	−4.9	.212	−5.5	<.001
Thiocyanate–plasma[b]	2.1	.750	3.0	.001	−13.5	.001	−7.6	<.001
CO[b]	−4.5	.441	2.6	.003	−8.5	.023	−6.8	<.001
COHb[b]	−6.2	.382	6.3	<.001	−10.6	.019	−5.0	<.001
Multivariate analyses								
Interviewer administered	4.0	.402	1.2	.066	3.8	.204	8.1	<.001
Student population	−5.2	.301	−3.3	.001	0.1	.983	−0.3	.551
Observational study	3.7	.480	0.8	.310	3.5	.286	5.2	<.001
Cotinine–saliva[b]	1.4	.845	2.9	<.001	−5.4	.222	−8.6	<.001
Thiocyanate–plasma[b]	2.5	.716	2.6	<.001	−12.3	.007	−10.7	<.001
CO[b]	−1.5	.809	2.9	<.001	−6.9	.085	−3.8	<.001
COHb[b]	−4.5	.556	5.6	<.001	−8.2	.097	−9.2	<.001

Note. Four studies using thiocyanate–saliva samples were omitted from analysis. CO = carbon monoxide; COHb = carboxyhemoglobin, measured in percent.
[a]Estimated percentage point difference in sensitivity or specificity due to the study characteristic.
[b]Cotinine–plasma is the reference.

TABLE 3—Estimated Effects of Study Characteristics on Sensitivity and Specificity of Self-Reported Smoking: Independent Comparisons Only (n = 30)

	Sensitivity				Specificity			
	Ordinary Least Squares Random Effects		Logistic: Fixed Effects		Ordinary Least Squares Random Effects		Logistic: Fixed Effects	
Study Characteristic	Effect[a]	*P*	Effect[a]	*P*	Effect[a]	*P*	Effect[a]	*P*
Bivariate analyses								
Interviewer administered	4.9	.326	2.4	.001	4.2	.056	6.2	<.001
Student population	−1.7	.351	−4.6	<.001	−1.5	.508	−4.7	<.001
Observational study	5.3	.389	2.1	.033	0.9	.748	1.3	.239
Cotinine–saliva[b]	0.1	.993	3.3	.010	−8.1	.026	−7.1	<.001
Thiocyanate–plasma[b]	2.5	.804	3.6	.017	−9.0	.031	−9.2	<.001
CO[b]	−4.5	.546	1.6	.128	−8.4	.007	−7.0	<.001
COHb[b]	−5.9	.511	8.0	<.001	−7.0	.051	−4.3	<.001
Multivariate analyses								
Interviewer administered	4.0	.554	2.9	.002	4.6	.088	6.7	<.001
Student population	−7.1	.364	−3.7	.003	2.3	.454	0.4	.469
Observational study	9.3	.272	8.7	<.001	1.3	.688	0.8	.406
Cotinine–saliva[b]	4.4	.667	3.9	<.001	−9.5	.022	−10.7	<.001
Thiocyanate–plasma[b]	5.3	.639	3.9	<.001	−6.9	.117	−8.9	<.001
CO[b]	1.2	.888	3.9	<.001	−7.4	.030	−5.1	<.001
COHb[b]	−0.9	.931	−8.5	<.001	−4.9	.245	−7.5	<.001

Note. Two studies using thiocyanate–saliva samples were omitted from analysis. CO = carbon monoxide; COHb = carboxyhemoglobin, measured in percent.
[a]Estimated percentage point difference in sensitivity or specificity due to the study characteristic.
[b]Cotinine–plasma is the reference.

sensitivity than cotinine–plasma, and interviewer-administered studies provided marginally greater sensitivity ($P = .047$).

For specificity, the random-effects model showed statistically significant lower estimates for thiocyanate–plasma, saliva cotinine, and carbon monoxide (both air and plasma) in comparison with plasma cotinine. When all study characteristics

were entered into the regression equation, significant beneficial effects for specificity remained only for cotinine–plasma in comparison with plasma thiocyanate samples.

Estimates of self-report accuracy using the fixed-effects analyses were much more likely to show statistically significant differences due to study and population characteristics than were the results of random-effects analyses. The ordinary least squares random-effects results in Table 2 (columns 2 and 4) illustrate the most conservative approach; the logistic fixed-effects results (columns 3 and 5) are less conservative under the assumption that studies are homogeneous. The ordinary least squares random-effects analyses, both bivariate and multivariate, showed no significant effects for sensitivity and significantly lower estimates of specificity for plasma thiocyanate samples using statistical significance as the criterion.

Table 3 shows the results of even more conservative analyses using only 30 independent comparisons, with 1 comparison chosen at random from each study eligible for analyses. Results were similar for sensitivity using the random-effects or the fixed-effects approach in both bivariate and multivariate analyses. No study characteristics produced significantly higher or lower estimates of self-report accuracy. Fixed-effects analyses of sensitivity were also similar for bivariate and multivariate analyses, indicating that all study characteristics produced the same higher or lower estimates. Results of studies with student populations yielded lower estimates of sensitivity. For specificity and the random effects analyses, significantly lower estimates were obtained for the different samples in comparison with plasma cotinine. All study characteristics were significantly different in the bivariate, fixed-effects analysis of specificity, except for observational studies (in comparison with intervention studies). This same pattern emerged with multivariate analyses of specificity using the fixed-effects approach, although the differences obtained from student populations no longer were evident.

An analysis of two-way interaction effects using the random-effects model and 47 comparisons yielded only one significant interaction each for sensitivity and specificity (interviewer-administered questionnaires in observational studies). Our power to detect interaction effects, however, was small given the small number of studies included in the analysis.

Discussion

This meta-analysis of published studies comparing self-reported smoking status with results of biochemical validation suggests generally high levels of sensitivity and specificity for self-report. Across all studies, the sensitivity of self-report was 87%, and the specificity was 89%. Nonetheless, both measures of accuracy proved quite variable among studies, as shown in Table 1, suggesting that specific aspects of the setting, study population, measurement methods, and study purpose are important to the accuracy of smoking self-reports.

Our search for systematic patterns of variation in sensitivity and specificity across studies was only partially successful. Two different methods of meta-analysis yielded generally similar results on the sign and magnitude of the effect of each study characteristic on sensitivity or specificity, but they often produced widely divergent verdicts on the statistical significance of those effects. The fact that the random-effects model produced *P* values that were usually much higher than those from the fixed-effects model indicates a large amount of between-study variability that could not be accounted for by the study characteristics measured. This conclusion, in turn, suggests that other unmeasured study characteristics may confound our results about the effects of method of administration, study population, study type, and biochemical test and specimen.

One of the most important unmeasured study characteristics is the specific wording of questions on smoking status. Very few studies reported this critical information, despite considerable evidence from survey research that responses are heavily influenced by how a question is phrased and the order in which questions are asked.

Because of these limitations, the observed patterns of association between accuracy of self-report and study characteristics must be interpreted with caution. The results suggest that interviewer-administered questionnaires yielded higher estimates of sensitivity and specificity than did self-administered questionnaires. Interviews identified more of the smokers correctly and classified nonsmokers more accurately. This may reflect smokers' awareness of sensory cues about their smoking (visible cigarettes, nicotine stains on teeth or hands, smoke odor) that would be obvious to an interviewer. More respondents may attempt to hide smoking behavior in self-administered questionnaires, even when biochemical validation is known.

Even in the most conservative analyses, student self-reports had lower sensitivity than studies using reports from subjects in the general population; however, the results were not always statistically significant. That is, students appear more likely to deny smoking, even when biochemical measures classify them as smokers. This is not surprising, since smoking by minors is illegal in most states and many young tobacco users have not yet defined themselves as smokers. Both of these conditions would contribute to a tendency, whether conscious (fear of being found out) or unconscious (self-definition inconsistent with behavior), to underreport. The different analyses do not suggest, however, that student self-reports have higher specificity than do those of subjects from the general population. Unfortunately, all studies of student populations reported here were observational in nature; no intervention studies with students that reported biochemical results were found among published studies, although the bogus pipeline procedures have been used with students. Reports of accuracy from intervention studies with student populations might have yielded lower estimates of sensitivity given the results from observational studies.

In the most conservative analyses, using only one comparison from each study and the random-effects approach, observational studies had higher levels of sensitivity than intervention studies, a conclusion supported by a qualitative review of studies in the 1990 surgeon general's report.[3] Self-reports from subjects in intervention studies, in which there is an expectation of cessation of smoking, are more likely to involve underreporting of actual smoking.

Self-reports of participants in intervention programs also have lower specificity, meaning that more biochemically validated nonsmokers reported smoking. Biochemical tests have limited ability to detect the very low levels of smoking that would be expected from recent quitters who "slip" and smoke an occasional cigarette.[34] Discussions of quitters' reactions to such slips (the abstinence violation effect) suggest that the would-be quitter is likely to exaggerate the importance of a few cigarettes under these conditions.[39] This can produce a report of "smoking," although its magnitude is smaller than the test can detect. In observational studies, with little focus on

cessation, this reaction may not be triggered.

Consistent with previous reports,[40] self-reports of smoking validated by means of cotinine–plasma biochemical measures appear to have higher specificity than those reports validated by other biochemical tests and specimens. The *P* values for tests comparing cotinine–plasma and other biochemical measures were not always significant in the more conservative random effects model, but the signs and magnitudes of the coefficients were mostly consistent across analytic models and between Tables 2 and 3. The observed pattern shows that more false-positive self-reports (i.e., respondents report that they are smokers when the biochemical test does not confirm such a report) tend to be observed in studies that use methods other than cotinine–plasma. This finding may reflect variation in the accuracy of the other biochemical tests in relation to self-reported smoking. These so-called false-positives may actually be smokers, but the poorer biochemical tests were too insensitive to detect relatively low levels of biochemical abnormality. There is no substantive reason to expect differences in self-report behavior due to the form of biochemical validation.

The fixed-effects and random-effects analyses yielded both different point estimates of the effects and different significance levels. Between-study variation was highly significant in all analyses, calling into question a key premise underlying the fixed-effects model. The results obtained from using multiple nonindependent comparisons in the different studies (Table 2) and from using only independent measures (Table 3) were substantially similar. Thus, the lack of independence in samples did not prove to be a serious problem for this meta-analysis.

In summary, our results suggest that biochemical validation may be more important in intervention studies, in studies with student populations, and in studies using self-administered rather than interviewer-administered questionnaires. For greatest accuracy, self-administered questionnaires given to students might benefit from biochemical validation, given the lower estimates obtained from these groups. Cotinine–plasma may be the biochemical test of choice if adequate resources are available for collection and analysis.

The decision to use biochemical validation is not as straightforward as it might appear. Biochemical validation is costly and sometimes difficult to obtain for all participants.[6] The bogus pipeline procedure and use of biochemical assessments with random subsamples of the target population are alternative strategies.[4]

The conclusions from this meta-analysis are subject to three major cautionary notes that, in turn, indicate needed improvements in the conduct and reporting of future studies of smoking behavior. First, it is known that the form and content of self-report questions about smoking influence the responses given and, hence, the categorization of respondents as smokers.[25] This observation argues that studies asking about smoking should report or reference the exact questions used so that this source of misclassification can be controlled. Second, authors of published studies need to make clear how biochemical validation was presented to study participants. In studies in which participants know that biochemical assessment will occurr, such as in the bogus pipeline procedure, self-reported smoking rates may be different from those in studies in which biochemical assessment is presented only at the point of collection. The exact procedures used in studies should be identified in the methods sections of articles to permit evaluation of the results according to the potential bias introduced by precollection announcement of biochemical validation.

Finally, in any meta-analysis, publication bias and the "file drawer phenomenon," the failure to submit for publication studies that do not produce effects, have an impact on the data available for analysis and can bias the outcome.[41,42] Although this bias is likely to be reduced in smoking studies that do not seek to test a specific hypothesis about validity, the universe of potential data sets on the validity of self-reports is still influenced by investigator analysis and submission of data for publication. This meta-analysis also excluded studies of pregnant women, and, thus, the generalizability of the results must be considered in comparison with the population under investigation.

The failure to include sufficient data on biochemical assessment resulted in the exclusion of several studies that reported validations of self-report. This suggests that in future studies with smoking validation, sufficient data should be published for investigators to confirm and evaluate the accuracy of self-reports. □

Acknowledgments

This work was supported by grant CA 34847 from the National Cancer Institute awarded to the Fred Hutchinson Cancer Center Public Health Sciences Division, Cancer Prevention Research Unit, for work on the methodology of community-based studies.

An earlier version of this paper was presented at the International Conference on Measurement Errors in Survey, Tucson, Ariz, November 1990.

References

1. *Reducing the Health Consequences of Smoking: 25 Years of Progress. A Report of the Surgeon General.* Washington, DC: US Dept of Health and Human Services; 1989. DHHS publication CDC 89-8411.
2. Haley NJ, Hoffmann D. Analysis for nicotine and cotinine in hair to determine cigarette smoker status. *Clin Chem.* 1985;31: 1598–1600.
3. *The Health Benefits of Smoking Cessation.* Washington, DC: US Dept of Health and Human Services; 1990. DHHS publication CDC 90-8416.
4. Murray DM, O'Connell CM, Schmidt LA, Perry CL. The validity of smoking self-reports by adolescents: a reexamination of the bogus pipeline procedure. *Addict Behav.* 1987;12:7–15.
5. Luepker RV, Pallonen UE, Murray DM, Pirie PL. Validity of telephone surveys in assessing cigarette smoking in young adults. *Am J Public Health.* 1989;79:202–204.
6. Velicer WF, Prochaska JO, Rossi JS, et al. Assessing outcome in smoking cessation studies. *Psychol Bull.* 1992;111:23–47.
7. Benowitz NL. The use of biologic fluid samples in assessing tobacco smoke consumption. In: Grubowski J, Bell CS, eds. *Measurement in the Analysis and Treatment of Smoking Behavior.* Washington, DC: US Dept of Health and Human Services; 1983;6–26. NIDA research monograph 48.
8. Windsor T, Morris J, Cutter G, et al. Sensitivity, specificity and predictive value of saliva thiocyanate among pregnant women. *Addict Behav.* 1989;14:447–452.
9. Light RJ, Pillemer DB. *Summing Up: The Science of Reviewing Research.* Cambridge, Mass: Harvard University Press; 1984.
10. Rosenthal R. *Meta-Analytic Procedures for Social Research.* Newbury Park, Calif: Sage; 1984. Applied Social Research Methods Series Volume 6.
11. Hunter JE, Schmidt FL. *Methods of Meta-Analysis.* Newbury Park, Calif: Sage; 1990.
12. Gianrossi R, Detrano R, Mulvihill D, et al. Exercise-induced ST depression in the diagnosis of coronary artery disease. A meta-analysis. *Circulation.* 1989;80:87–98.
13. Phillips KA. The use of meta-analysis in technology assessment: a meta-analysis of the enzyme immunosorbent assay human immunodeficiency virus antibody test. *J Clin Epidemiol.* 1991;44:925–931.
14. Bauman KE, Dent CW. Influence of an objective measure on self-reports of behavior. *J Appl Psychol.* 1982;67:623–628.
15. Bauman KE, Koch GG, Bryan E. Validity of self-reports of adolescent cigarette smoking. *Int J Addict.* 1982;17:1131–1136.
16. Bauman KE, Koch GG. Validity of self-

reports and descriptive and analytical conclusions: the case of cigarette smoking by adolescents and their mothers. *Am J Epidemiol.* 1983;118:90–98.
17. Cohen JD, Bartsch GE. A comparison between carboxyhemoglobin and serum thiocyanate determinations as indicators of cigarette smoking. *Am J Public Health.* 1980;70:284–286.
18. Fortmann SP, Rogers T, Vranizan K, et al. Indirect measures of cigarette use: expired-air carbon monoxide versus plasma thiocyanate. *Prev Med.* 1984;13:127–135.
19. Gillies PA, Wilcox B, Coates C, et al. Use of objective measurement in the validation of self-reported smoking in children aged 10 and 11 years: saliva thiocyanate. *J Epidemiol Community Health.* 1982;36:205–208.
20. Haddow JE, Palomakii GE, Knight GJ. Use of serum cotinine to assess the accuracy of self-reported non-smoking. *BMJ.* 1986;293:1306. Letter.
21. Luepker RV, Pechacek TF, Murray DM, et al. Saliva thiocyanate: a chemical indicator of cigarette smoking in adolescents. *Am J Public Health.* 1981;71:1320–1324.
22. McNeill AD, Jarvis MJ, West R, et al. Saliva cotinine as an indicator of cigarette smoking in adolescents. *Br J Addict.* 1987;82: 1355–1360.
23. Noland MP, Kryscio RJ, Riggs RS, et al. Saliva cotinine and thiocyanate: chemical indicators of smokeless tobacco and cigarette use in adolescents. *J Behav Med.* 1988;11:423–433.
24. Pechacek TF, Murray DM, Luepker RV, et al. Measurement of adolescent smoking behavior: rationale and methods. *J Behav Med.* 1984;7:123–130.
25. Petitti DB, Friedman GD, Kahn W. Accuracy of information on smoking habits provided on self-administered research questionnaires. *Am J Public Health.* 1981; 71:308–311.
26. Pierce JP, Dwyer T, DiGiusto E, et al. Cotinine validation of self-reported smoking in commercially run community surveys. *J Chronic Dis.* 1987;40:689–695.
27. Pojer R, Whitefield JB, Poulos V, et al. Carboxyhemoglobin, cotinine, and thiocyanate assay compared for distinguishing smokers from non-smokers. *Clin Chem.* 1984;30:1377–1380.
28. Prignot J. Quantification and chemical markers of tobacco-exposure. *Eur J Respir Dis.* 1987;70:1–7.
29. Saloojee Y, Vesey CJ, Cole PV. Carboxyhemoglobin and plasma thiocyanate: complementary indicators of smoking behavior? *Thorax.* 1982;37:521–525.
30. Sillett RW, Wilson MB, Malcolm RE, et al. Deception among smokers. *BMJ.* 1978;2: 1185–1186.
31. Slattery ML, Hunt SC, French TK, et al. Validity of cigarette smoking habits in three epidemiologic studies in Utah. *Prev Med.* 1989;18:11–19.
32. Stookey GK, Katz BP, Olson BI, et al. Evaluation of biochemical validation measures in determination of smoking status. *J Dent Res.* 1987;66:1597–1601.
33. Van Vunakis H, Tashkin DP, Rigas B, et al. Relative sensitivity and specificity of salivary and serum cotinine in identifying tobacco-smoking status of self-reported nonsmokers and smokers of tobacco and/or marijuana. *Arch Environ Health.* 1989;44: 53–58.
34. Vogt TM, Selvin S, Widdowson G, et al. Expired air carbon monoxide and serum thiocyanate as objective measures of cigarette exposure. *Am J Public Health.* 1977;67: 545–549.
35. Wagenknecht LE, Cutter GR, Haley NJ, et al. Racial differences in serum cotinine levels among smokers in the coronary artery risk development in (young) adults study. *Am J Public Health.* 1990;80:1053–1056.
36. Wald NJ, Idle M, Boreham J, et al. Carbon monoxide in breath in relation to smoking and carboxyhemoglobin levels. *Thorax.* 1981;36:366–369.
37. Williams CL, Eng A, Botwin GJ, et al. Validation of students' self-reported cigarette smoking levels with plasma cotinine levels. *Am J Public Health.* 1979;69:1272–1274.
38. Laird NM, Mosteller F. Some statistical methods for combining experimental results. *Int J Technol Assess Health Care.* 1990;6:5–30.
39. Curry SG, Marlatt GA, Gordon JR. Abstinence violation effect: validation of an attributional construct with smoking cessation. *J Consult Clin Psychol.* 1987;55:145–149.
40. Murray RP, Connett JE, Lauger GG, Voelker HT. Error in smoking measures: effects of intervention on relations of cotinine and CO to self-reported smoking. *Am J Public Health.* 1993;83:1251–1257.
41. Jenicek M. Meta-analysis in medicine: where we are and where we want to go. *J Clin Epidemiol.* 1989;42:35–44.
42. Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J Clin Epidemiol.* 1991;44:127–139.

Estimating the Number of HIV-Infected Injection Drug Users in Bangkok: A Capture–Recapture Method

ABSTRACT

Objectives. The purpose of the study was to estimate the number of injection drug users infected with the human immunodeficiency virus (HIV) in Bangkok to allow planning for health services for this population.

Methods. A two-sample capture–recapture method was used. The first capture listed all persons on methadone treatment for opiate addiction from April 17 through May 17, 1991, at 18 facilities in Bangkok. The second capture involved urine testing of persons held at 72 Bangkok police stations from June 3 through September 30, 1991. Persons whose urine tests were positive for opiate metabolites or methadone were included on the second list.

Results. The first capture comprised 4064 persons and the recapture 1540 persons. There were 171 persons included on both lists, yielding an estimate of 36 600 opiate users in Bangkok. Existing data indicate that 89% of opiate users in Bangkok inject drugs and that about one third are infected with HIV, yielding an estimate of approximately 12 000 HIV-infected injection drug users in Bangkok in 1991.

Conclusions. During the 1990s the number of cases of acquired immunodeficiency syndrome (AIDS) and other HIV-related diseases, including tuberculosis, in the population of HIV-infected injection drug users in Bangkok will increase dramatically, placing new demands on existing health care facilities. The capture–recapture method may be useful in estimating difficult-to-count populations, including injection drug users. (*Am J Public Health.* 1994;84: 1094–1099)

Timothy D. Mastro, MD, Dwip Kitayaporn, MD, DrPH, Bruce G. Weniger, MD, MPH, Suphak Vanichseni, MD, MPH, PhD, Vuti Laosunthorn, PhD, Thongchai Uneklabh, MD, Chintra Uneklabh, MD, Kachit Choopanya, MD, MPH, and Khanchit Limpakarnjanarat, MD, MPH

Introduction

The explosive initial phase of the epidemic of human immunodeficiency virus type 1 (HIV-1) infection in Thailand occurred in 1988 among injection drug users in Bangkok, the capital, presumably as a result of transmission via the sharing of injection equipment.[1,2] The rate of HIV-1 seropositivity among opiate users seeking drug treatment in this city of 6 million people rose from 0% to 1% in late 1987 to 32% to 43% in late 1988.[2] From 1989 through 1993, seroprevalence among persons in opiate treatment programs stabilized at approximately 35% to 40%.[2] Existing data indicate that heroin is the principal drug used by 97% of persons seeking drug treatment in Bangkok and that 89% of such persons inject the drug.[3] Since 1988 the HIV-1 epidemic in Thailand has extended to persons with other risk behaviors for HIV infection; rates of HIV-1 seropositivity have increased dramatically among female prostitutes, young men seeking treatment at sexually transmitted disease clinics, and 21-year-old male military conscripts.[2,4]

An accurate estimate of the total number of injection drug users in Bangkok could be used to estimate the number of HIV-infected injection drug users. This estimate would be helpful in assessing the amount of HIV transmission that is due to injection drug use as compared with transmission due to sexual activity. Such an estimate could also be used to plan drug treatment programs and other health services for this population. In particular, an estimate of the number of HIV-infected injection drug users would allow for a reasonable prediction of the number of acquired immunodeficiency syndrome (AIDS) cases to expect among this population in the coming decade. The care of injection drug users with AIDS and HIV-related infections such as tuberculosis will likely place a great burden on existing health care facilities in Bangkok and elsewhere in Thailand.

We employed a capture–recapture method to estimate the number of opiate users in Bangkok and then used existing data to estimate the total number of injection drug users and the number of HIV-infected injection drug users in this city in 1991.

Background

Opiate use in Thailand dates back more than 300 years. Although there were

Timothy D. Mastro, Dwip Kitayaporn, Bruce G. Weniger, and Khanchit Limpakarnjanarat are with the HIV/AIDS Collaboration, Bangkok, Thailand. Dr Mastro and Dr Weniger are also with the National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Ga. Dr Kitayaporn is also with the Faculty of Tropical Medicine, Mahidol University, Bangkok. Suphak Vanichseni and Kachit Choopanya are with the Department of Health, Bangkok Metropolitan Administration. Vuti Laosunthorn is with the Office of the Narcotics Control Board, Bangkok. Thongchai Uneklabh and Chintra Uneklabh are with Thanyarak Hospital, Ministry of Public Health, Pathum Thani, Thailand.

Requests for reprints should be sent to Timothy D. Mastro, MD, HIV/AIDS Collaboration, 88/7 Soi Bamrasnaradura, Tivanon Rd, Nonthaburi 11000, Thailand.

This paper was accepted December 28, 1993.

Note. The use of trade names is for identification only and does not imply endorsement by the Public Health Service or by the US Department of Health and Human Services.

Editor's Note. See related annotation by Neugebauer and Wittes (p 1068) in this issue.

laws in the 18th century prohibiting the use of opium, the trade in and use of opium were regulated until 1959, when opium use was banned.[5] In 1959, opium smokers were required to register, resulting in the documentation of 70 985 users, about half of whom were in Bangkok.[5] Heroin use was not noted until shortly after 1959, and the practice of injecting heroin, predominantly by young men, increased in the second half of the 1960s.[5] Opium poppies have traditionally been grown in the mountainous Golden Triangle region that includes northern Thailand, Myanmar (Burma), and Laos. This region is one of the world's foremost heroin-producing areas. The street price (200–500 baht [US$8–20] per gram) of 60% to 90% pure ("No. 4 grade") heroin in Bangkok, 700 km to the south, is among the lowest in the world. Compared with the official daily minimum wage of 100 baht (US$4) in Bangkok, the relative cost of regular heroin use is considerably lower in Thailand than in other countries where illicit heroin use is a problem. A recent study in Bangkok indicated that 69% of 600 injection drug users had some form of legitimate employment, but also that 69% had spent time in prison since they began injecting heroin.[1]

Free treatment services for drug users in Bangkok are provided by the Bangkok Metropolitan Administration and the Ministry of Public Health. Law enforcement control is waived for clients.[6] The Bangkok Metropolitan Administration operates 17 outpatient drug treatment clinics throughout the city and the Ministry of Public Health operates Thanyarak Hospital, a 500-bed inpatient facility with an active outpatient department, in Pathum Thani on the northern edge of Bangkok. Methadone treatment, usually lasting 45 days, is the principal mode of opiate detoxification. Routine urine screening for opiates indicates that a large proportion (40%–80%) of persons receiving outpatient methadone treatment continue to use heroin while in treatment (Bangkok Metropolitan Administration, unpublished data). Admissions to treatment are reported to the Drug Dependence Information System, operated by the Office of the Narcotics Control Board of the Office of the Prime Minister and the Department of Medical Services, Ministry of Public Health, in collaboration with the Institute of Health Research, Chulalongkorn University.[6] In 1989, this system recorded 28 457 courses of drug treatment in Bangkok; 95% of the clients were men, 87% were aged 20 through 39 years, and 82% were outpatients.[3] Multiple treatments for the same person were recorded as separate events.

The total number of injection drug users in Bangkok is not reliably known; however, efforts have been made to estimate the number of persons served by the 17 Bangkok Metropolitan Administration treatment facilities. For the years 1986 through 1989, the number of treatment admissions ranged from 22 522 to 24 355. When duplicate names were eliminated from admission records, the number of persons treated annually ranged from 7238 to 8591, indicating an average of three treatment courses annually per patient served.[7,8] These data were used in a multiple-recapture analysis (using admission data from the three trimesters of a year as separate captures) to estimate the drug treatment population. In this analysis, the treatment population in the years 1986 through 1989 was estimated to be between 16 460 and 21 655 persons.[7,8] However, this method was thought to underestimate the total opiate-using population in Bangkok because it did not adequately account for persons not receiving treatment.

In 1989, the Bangkok Metropolitan Administration conducted a modified key informant study[9] of persons receiving methadone treatment to determine the proportion of heroin users in treatment. Clients were asked to list the persons they knew who used heroin and to indicate whether these persons had ever been admitted to a treatment program. The mean number of persons listed on 1944 questionnaires was 5.3 (SD = 3.2). The proportion of heroin users who had ever been in treatment was 0.626.[10] Dividing the estimated annual treatment population (16 460 to 21 655) by this proportion yields an estimate of 26 294 to 34 593 total heroin users. However, because this method assumes that all heroin users are known by drug users in treatment, it may underestimate the actual heroin-using population.

Methods

The study protocol was approved by the Ethical Review of Research Committee of the Thai Ministry of Public Health. A two-sample capture–recapture method was used.[11,12] The first capture was of all persons who were enrolled in methadone treatment programs for opiate addiction during the period April 17 through May 17, 1991, at the 17 clinics and at Thanyarak Hospital. For Thanyarak Hospital, only Bangkok residents and only those on outpatient treatment or those discharged from inpatient treatment were included. A confidential list was compiled containing forename and surname, gender, age, date of birth, Thai national identification number, and date of last methadone treatment.

The recapture was conducted from June 3 through September 30, 1991. Study teams made 891 visits on 66 days to 72 police stations in all areas of Bangkok. Included in the survey were persons held at police stations who had been arrested within the previous 72 hours. Urine samples were collected and the following information was recorded confidentially: forename and surname, gender, age, date of birth, Thai national identification number, and reason for arrest. Urine was tested by enzyme immunoassay for the presence of opiate metabolites (Emit d.a.u. Opiate Assay, Syva Co, Palo Alto, Calif) and methadone (Emit d.a.u. Methadone Assay, Syva Co, Palo Alto, Calif). The cutoff for positive, 300 ng/mL, was recommended by the manufacturer and was described as being 100% predictive of the presence of these substances. Information on persons with urine tests positive for opiate metabolites or methadone was included on a list for further analysis.

Information collected in the lists from the two captures was entered twice into a computer database in the Thai language. After keypunch errors were corrected, duplicate entries were removed from each list by means of a combination of computer sorting and visual inspection. The two lists were then merged, computer sorted by multiple variables, and visually inspected to identify the same entries (matches) in both lists. Criteria used to determine matching entries are shown in Table 1. Names were considered similar if they contained only minor spelling differences that did not change the name. This definition was established in view of the complexity of the Thai alphabet, which consists of 44 consonants, 32 simple and compound vowels (some written as super- or subscripts or combinations), and 4 tone markers. Dates of birth were considered similar if they varied by no more than one digit.

An estimate of the total population of opiate users (n) was made with the formula:

$$\text{Estimate of } n = \hat{n} = \frac{c_1 c_2}{m},$$

where c_1 and c_2 are the numbers of persons in the first and second capture,

TABLE 1—Criteria Used to Determine Matching Persons in the Two Captures and Number of Matches Identified for Each Match Type

Match Type	No. of Matches	Forename	Surname	Age	Date of Birth	Thai ID Number
A	9	Same	Same	Same	Same	Same
B	65	Same	Same	Same	Same	...
C	52	Same	Same	Same	Similar	...
D	45	Similar/same	Similar/same	Same	Similar/same	...
E	4	Different	Similar/same	Same	Similar	...
F	15	Similar/same	Similar/same	Different	Different	...

Note. Gender was the same for all match types.

TABLE 2—Reasons for Arrest for Persons Included in the Second (Police Arrest) Capture (n = 1540)

Reason for Arrest	No.	%
Marijuana-related	833	54.1
Heroin-related	421	27.3
Gambling	99	6.4
Robbery	73	4.7
Inhalant use	49	3.2
Assault	10	0.6
Other	55	3.6

respectively, and m is the number of persons identified in both captures (matches).[13] The following formulas were used to calculate the variance (Var) and 95% confidence interval (CI) for the estimate of n:

$$\text{Var}(\hat{n}) = \frac{c_1 c_2 (c_1 - m)(c_2 - m)}{m^3}$$

$$95\%\ \text{CI} = \hat{n} \pm 1.96\sqrt{\text{Var}(\hat{n})}.$$

These formulas[13] were also used to determine age-stratum–specific estimates of n. Stratum-specific estimates of n and variances were then summed to determine the total n and 95% confidence interval.

Population rates were calculated using 6 million as the total Bangkok population and 1 244 000 (20.7%) as the number of 20- through 39-year-old male Bangkok residents.[14]

Results

In the first capture, 4064 persons remained on the list after 478 duplicate names were removed. The mean age was 32.5 years (SD = 7.7); 34.0% were aged 20 through 29 years; 52.0% were aged 30 through 39 years; and 93.6% were male. Name and gender were recorded for all persons; also recorded were age for 99.7%, date of birth for 88.3%, and national identification number for 79.9%.

In the recapture, 8212 persons were surveyed at police stations. Included on the list (following the removal of 105 duplicate names) were the 1540 (18.8%) persons whose urine tests were positive for opiate metabolites (1513 [98.2%]) or methadone (198 [12.9%]); 171 (11.1%) were positive on both assays. The mean age of those with positive urine tests was 29.8 years (SD = 7.0); 45.4% were aged 20 through 29 years; 41.8% were aged 30 through 39 years; and 94.3% were male. Name and gender were recorded for all persons; also recorded were age for 99.9%, date of birth for 94.3%, and national identification number for only 2.8%. The reasons for arrest are shown in Table 2. Reportedly, persons possessing heroin were commonly charged with marijuana possession to expedite processing at police stations.

Match types A through D (Table 1) were considered to be true matches. By this standard, comparisons of the two lists identified 171 matches, yielding an estimate of 36 600 (95% CI = 31 538, 41 662) opiate users in Bangkok in 1991 (Table 3). Since the age distributions in the two captures were different ($P < .01$), a stratified analysis by age groups was conducted, yielding an estimate of 36 376 (95% CI = 30 996, 41 757) opiate users.

Various criteria were used to estimate n (Table 4). Using more restrictive (more specific) matching criteria by excluding type D matches increases the estimate of n by 35.7%, to 49 671. Conversely, using less restrictive (more sensitive) criteria by including type E and F matches decreases the estimate of n by 10.0%, to 32 940.

An estimate of 32 574 injection drug users (5 per 1000 persons in the total Bangkok population) results from the use of existing data indicating that 89% of opiate users seeking drug treatment in Bangkok inject the drugs.[1] However, since 94% of injection drug users are men and 87% are aged 20 through 39 years, the rate of injection drug users among men aged 20 through 39 years is 21 per 1000. Data from drug treatment programs indicate that roughly one third of all opiate users were HIV-1 seropositive in 1991[2]; it is likely that all were injection drug users, resulting in an estimate of approximately 12 000 injection drug users infected with HIV-1 (2 per 1000 total population). The rate of HIV-infected injection drug users among men aged 20 through 39 years is 8 per 1000.

Discussion

In 1991, it was estimated that there were 300 000 HIV-infected persons in Thailand.[2] Epidemiologic evidence suggested that the majority of these infections resulted from heterosexual transmission[2]; however, limited data were available to estimate the number of HIV-infected injection drug users. Capture–recapture methodology can be used to estimate the size of difficult-to-count subpopulations, although the reliable use of this method is not easily achieved.[15,16] It was possible to use this method in Bangkok, in part, owing to the availability of drug treatment and the great variety of Thai names.

Interest in enumerating injection drug user and prostitute populations, because of their increased risk for HIV transmission, recently led the World Health Organization's Global Programme on AIDS to review capture–recapture methods.[17] These methods have been used extensively in biologic sciences to enumerate animal populations[18,19] and have been used in the United States to evaluate the completeness of national reporting systems for congenital rubella syndrome[20] and mortality due to tetanus[21] and pertussis.[22] The numbers of drug users in Amsterdam,[23] in two boroughs of London,[24] and in Rome[25] have been estimated by a two-capture method. More recently, a multiple-recapture method was used to estimate the numbers of injection drug users[26–28] and female street prostitutes[29] in Glasgow, Scotland. Four major assumptions must be satisfied for the two-capture method to produce reliable results.[11,12,30]

The first assumption is that the population is closed. In reality, no city is closed and this assumption can only be

satisfied to a reasonable degree. If the population of opiate users is very dynamic, with members entering and leaving with great frequency, the recapture will be less likely to include persons identified in the first capture, resulting in a falsely low number of matches and an overestimate of the population of opiate users. To allow for the assumption that the opiate user population was reasonably stable during the study period, we limited the duration of the study to less than 6 months. Although we could not determine the number of opiate users who entered or left the city, it is unlikely that a substantial proportion did so during this brief period.

The second assumption is that the capture sources are independent. There is no clear positive or negative dependence between receiving methadone treatment and, at a later time, being arrested. The two captures were conducted sequentially with a 2-week period between them to allow for persons receiving methadone treatment in April or May to be at risk for arrest in June through September. The urine testing employed in the recapture permitted the detection of persons from the first capture who had either continued methadone or stopped methadone while using opiates. It could be argued that persons receiving methadone treatment have a decreased need for heroin (and therefore money) and are therefore less likely to be involved in criminal activity that would lead to arrest (negative dependence). However, in the second (police arrest) capture, we identified 198 persons with urine tests positive for methadone, 171 (86.4%) of whom were also positive for opiate metabolites. Conversely, one could argue that persons who receive methadone treatment are more likely to be arrested because of their increased likelihood of being identified as drug users by the police (positive dependence); however, we found no evidence that this occurred. The assumption of independence among captures is not required if more than two captures are employed; log-linear modeling is used to account for dependence among the captures.[13] In Bangkok, we were unable to identify a practical source for a third capture.

The third assumption is that all members of the population have the same probability of being captured. Free methadone treatment is available to all opiate users in Bangkok. All opiate users have a probability of seeking drug treatment and therefore of having been identified in the first capture; however, it is likely that there is a subgroup of opiate users, perhaps defined by socioeconomic status, with a decreased probability of seeking treatment. All opiate users have an appreciable probability of being arrested because heroin use in Thailand is illegal and highly associated with prison history.[1] The extensive scope of the police arrest recapture ensured a wide sampling of arrested persons in the city. Our finding that 34.0% of persons included in the methadone treatment capture, compared with 45.4% in the police station capture, were 20 through 29 years old suggests a difference in sampling for the two captures. Although an analysis stratified by age group yielded an estimate very similar to the overall estimate of the number of opiate users (Table 3), it is possible that other important strata remain unidentified and may contribute to a biased estimate.

The fourth assumption is that the capture history of each member is accurate: all true matches and only true matches are identified. The use of urine testing in the recapture ensured a highly specific identification of recent opiate users. Matches from the two capture lists were identified primarily by name, age, and date of birth. Gender was male for 94% of persons on both lists, and Thai national identification number, which was available for 79.9% of the persons on the first capture list but only 2.8% of those on the second list, was of limited value in matching. The great diversity of Thai names contributed to the specificity of the matching. Thai surnames, in the large majority of cases, are unique to each family and consequently are highly varied; thus it is extremely rare for two people to share the same forename and surname. Names reported to drug treatment programs are obtained from national identification cards and are therefore very accurate. The names reported at police stations generally are not verified by identification cards but are found to be largely reliable.

This study generated an estimate of the number of active opiate users in Bangkok at one point in time. Because we did not address the duration of injection drug use, we cannot estimate the number of persons in Bangkok who had ever injected drugs but who were not currently using heroin. However, a recent study of injection drug users in treatment in Bangkok found the average duration of use to be more than 8 years; only 11% had

TABLE 3—Estimates of the Number of Opiate Users in Bangkok (n), by Unstratified and Age-Stratified Analyses

Age Group	Capture 1	Capture 2	No. of Matches	n	95% Confidence Interval
Unstratified analysis					
All	4064	1540	171	36 600	31 538, 41 662
Stratified analysis					
<20 y	76	78	4	1482	105, 2859
20–29 y	1378	699	76	12 674	10 059, 15 289
30–39 y	2106	644	80	16 953	13 543, 20 363
>39 y	491	118	11	5267	2336, 8198
All	4051	1539	171	36 376	30 996, 41 757

TABLE 4—Estimates of the Number of Opiate Users in Bangkok (n) Based on the Number of Matches Identified[a]

Match Types	No. of Matches	n	95% Confidence Interval
A–B	74	84 575	65 946, 103 205
A–C	126	49 671	41 490, 57 852
A–D	171	36 600	31 538, 41 662
A–E	175	35 763	30 883, 40 643
A–F	190	32 940	28 658, 37 221

[a]Matches were determined according to the criteria shown in Table 1.

started injecting within the previous 2 years.[1] Because extensive HIV-1 transmission among injection drug users in Bangkok began only in 1988, these data indicate that most injection drug users infected with HIV are included in our estimate of 12 000, based on the estimated population of 36 600 active opiate users.

HIV-1 seroprevalence among opiate users in Bangkok has been measured only among those who seek treatment. Insufficient data are available to determine whether HIV-1 seroprevalence among injection drug users seeking treatment in Bangkok differs from the rate among those who do not seek treatment. Nevertheless, as a substantial number of injection drug users enter the treatment population each year, the stability of HIV-1 seroprevalence documented from 1988 through 1993 supports the assumption that such a difference is unlikely to be great.

We estimate that there were 36 600 active opiate users in Bangkok in 1991, leading to an estimate of approximately 12 000 HIV-infected injection drug users, including 0.8% of all men aged 20 through 39 years. Through 1992, 1675 cases of AIDS had been reported in Thailand (population 55 million) via a well-established surveillance system; 1150 cases (69%) were reported in 1992 alone, indicating a rapid increase in AIDS cases. In 1992, injection drug users accounted for 103 (9%) of new AIDS cases (Division of Epidemiology, Thai Ministry of Public Health, unpublished data). In the coming years, the number of cases of AIDS and other HIV-related diseases, including tuberculosis, arising from the present HIV-infected population of injection drug users will rise dramatically, placing new demands on existing health care facilities. Controlling HIV transmission among members of this population and their sexual partners will be important in limiting the AIDS epidemic in Bangkok. □

Acknowledgments

This paper was presented in part at the Eighth International Conference on AIDS/Third Sexually Transmitted Disease World Congress, July 19–24, 1992, Amsterdam, the Netherlands.

We thank Dr Roel A. Coutinho, Dr J. Arthur Woodward, Dr Vichai Poshyachinda, Dr Robert Byers, and Mark VanLandingham for their guidance in developing this study and Drs Timothy J. Dondero, Phillip Nieburg, and Glen A. Satten for reviewing the manuscript. We are grateful to Police Lt Gen Chavalit Yodmani, Dr Suchint Phalakornkule, Dr Sithaman Muttamara, Dr Kowit Wongpanich, Dr Suwat Chanchamnong, Dr Sudchai Panyarachun, Dr Thamnoon Vaniyapong, Dr Prawat Pisesskolkit, Dr Virasak Vachatith, Dr Wonchat Subhaturas, Dr Sittichat Puttiprasert, Dr Jidbhong Jayavasu, and Dr Pralom Sakuntanaga for their invaluable support. We also thank Siree Bunnag, M.L. Araya Sirodom, Vichet Puthaviriyakorn, Warunyupa Suwannadath, Sathit Sujarita, Busakorn Wongsuwan, Dr Pongvipa Lohsomboom, Dr Boondee Atikij, Baranee Mingmuang, Janjao Witta, and Mancharee Junk for their assistance in conducting this study. In addition, we are indebted to the staff of the Department of Health, Bangkok Metropolitan Administration, and the Technical and Research Section, Office of the Narcotics Control Board, for data and specimen collection; the laboratory staff of Thanyarak Hospital for specimen testing; and the personnel of the Bangkok Police Department for their cooperation.

We are grateful to Albert L. Bryant and Robert D. Griffiths and acknowledge the Narcotics Affairs Section of the US Embassy in Bangkok for financial support toward the cost of urine testing reagents.

References

1. Choopanya K, Vanichseni S, Des Jarlais DC, et al. Risk factors and HIV seropositivity among injecting drug users in Bangkok. *AIDS.* 1991;5:1509–1513.
2. Weniger BG, Limpakarnjanarat K, Ungchusak K, et al. The epidemiology of HIV infection and AIDS in Thailand. *AIDS.* 1991;5(suppl 2):S71–S85.
3. *Statistical Report FY 1989 of the Drug Dependence Information System: Treatment Population.* Bangkok, Thailand: Ministry of Public Health, Department of Medical Services, and Chulalongkorn University, Institute of Health Research; 1990.
4. Nopkesorn T, Mastro TD, Sangkharomya S, et al. HIV-1 infection in young men in northern Thailand. *AIDS.* 1993;7:1233–1239.
5. Poshyachinda V. *Heroin in Thailand.* Bangkok, Thailand: Chulalongkorn University, Institute of Health Research, Drug Dependence Research Center; February 1982. Technical report no. DD-1/82.
6. Poshyachinda V. Thailand data base on drug dependence and illicit drugs: an appraisal of merit for problem monitoring and intervention planning for demand reduction. In: Klein M, Sapienza H, McClain H Jr, Khan I, eds. *Clandestinely Produced Drugs, Analogues and Precursors: Problems and Solutions.* Washington, DC: US Dept of Justice, Drug Enforcement Administration; 1989.
7. Woodward JA. *Pilot Study: Estimating the Size of the Heroin Using Population Served by the 17 Clinics of the Bangkok Metropolitan Health Department.* Los Angeles, Calif: University of California; May 31, 1989. USAID Fixed Price Purchase Order 3316-991340.
8. Woodward JA. Design of a study to estimate the size of the heroin using population in Thailand. Los Angeles, Calif: University of California; October 1990.
9. Miller JD. The nominative technique: a new method of estimating heroin prevalence. In: Rouse BA, Kozel NJ, Richards LC, eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity.* Washington, DC: US Dept of Health and Human Services; 1985:158–171. National Institute of Drug Abuse Research Monograph 57.
10. Vanichseni S, Wongsuwan B, Woodward JA. *Modified Key Informant Estimation of the Proportion of IVDU in Treatment.* Bangkok, Thailand: Bangkok Metropolitan Administration; 1989.
11. Chandra Sekar C, Deming WE. On a method of estimating birth and death rates and the extent of registration. *Am Stat Assoc J.* 1949;44:101–115.
12. Wittes JT, Sidel VW. A generalization of the simple capture–recapture model with applications to epidemiological research. *J Chron Dis.* 1968;21:287–301.
13. Woodward JA, Retka RL, Ng L. Construct validity of heroin abuse estimators. *Int J Addict.* 1984;19:93–117.
14. *Thailand Health Profile.* Bangkok, Thailand: Ministry of Public Health; 1990.
15. Hook EB, Regal RR. Capture–recapture methods. *Lancet.* 1992;339:742. Letter.
16. McCarty DJ, Tull ES, Moy CS, Kwoh CK, LaPorte RE. Ascertainment corrected rates: applications of capture–recapture methods. *Int J Epidemiol.* 1993;22:559–565.
17. Taylor R. *A Review of Methods for Estimating the Size of Subgroups Particularly At Risk of Infection with HIV and Development of Proposals Which Could Be Used to Enumerate These Populations in the Field: With Particular Reference to the Use of Capture–Release Methods for Estimating Populations of IV Drug Users and Prostitutes.* Geneva, Switzerland: World Health Organization, Global Programme on AIDS; December 1989.
18. Cook LM, Brower LP, Croze HJ. The accuracy of population estimation from multiple recapture data. *J Animal Ecol.* 1967;36:57–60.
19. LaPorte RE, McCarty DJ, Tull ES, Tajima N. Counting birds, bees, and NCDs. *Lancet.* 1992;339:494–495. Letter.
20. Cochi SL, Edmonds LE, Dyer K, et al. Congenital rubella syndrome in the United States, 1970–1985. *Am J Epidemiol.* 1989; 129:349–361.
21. Sutter RW, Cochi SL, Brink EW, Sirotkin BI. Assessment of vital statistics and surveillance data for monitoring tetanus mortality, United States, 1979–1984. *Am J Epidemiol.* 1990;131:132–142.
22. Sutter RW, Cochi SL. Pertussis hospitalization and mortality in the United States, 1985–1988: evaluation of the completeness of national reporting. *JAMA.* 1992;267:386–391.
23. van Haastrecht HJA, van den Hoek JAR, Bardoux C, Leentvaar-Kuypers A, Coutinho RA. The course of the HIV epidemic among intravenous drug users in Amsterdam, the Netherlands. *Am J Public Health.* 1991;81:59–62.
24. Hartnoll R, Lewis R, Mitcheson M, Bryer S. Estimating the prevalence of opioid dependence. *Lancet.* 1985;i:203–205.
25. Davoli M, Arca M, Spadea T, et al. Estimating the number of drug injectors, a crucial step in predicting the HIV/AIDS incidence. Presented at the Eighth International Conference on AIDS/Third Sexually Transmitted Disease World Congress; July 19–24, 1992; Amsterdam, the Netherlands. Abstract PoC 4774.

26. Frischer M, Bloor M, Finlay A, et al. A new method of estimating prevalence of injecting drug use in an urban population: results from a Scottish city. *Int J Epidemiol.* 1991;20:997–1000.
27. Frischer M. Estimating prevalence of injecting drug use in Glasgow. *Br J Addict.* 1992;87:235–243.
28. Frischer M, Leyland A. Reliability of population and prevalence estimates. *Lancet.* 1992;339:995. Letter.
29. Bloor M, Leyland A, Barnard M, McKeganey N. Estimating hidden populations: a new method of calculating the prevalence of drug-injecting and non-injecting female street prostitutes. *Br J Addict.* 1991;86:1477–1483.
30. Woodward JA, Bonett DG, Brecht ML. Estimating the size of a heroin-abusing population using multiple-recapture census. In: Rouse BA, Kozel NJ, Richards LC, eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity.* National Institute of Drug Abuse Research Monograph 57. Washington, DC: US Dept of Health and Human Services; 1985;158–171.

Call for Abstracts for Epidemiology Late-Breaker Sessions

Oral Exchange Session

The Epidemiology Section will sponsor a late-breaker epidemiology oral exchange session on Wednesday, November 2, 1994, during the American Public Health Association's annual meeting October 30 through November 3 in Washington, DC. The exchange will provide a forum for the oral presentation of investigations, analyses, or methods that have been conceived, conducted, and/or completed so recently that authors could not meet the deadline (February 10, 1994) for regular submission to other epidemiology sessions.

Abstracts of fewer than 200 words (any format) and a stamped, self-addressed return envelope should be submitted to Polly A. Marchbanks, PhD, Chief, EIS Program, Centers for Disease Control and Prevention, Mailstop C-08, 1600 Clifton Rd, Atlanta, GA 30333; (404) 639-3588.

Abstracts must be *received* by September 12, 1994. Decisions will be made by September 19, 1994.

Poster Session

The Epidemiology Section will again sponsor a late-breaker poster session on Wednesday, November 2, 1994, at the APHA annual meeting October 30 through November 3 in Washington, DC. This session permits the presentation of work that has been completed too late in the last year for regular paper submission. Abstracts should report on work conducted during the last 6 to 12 months.

With a stamped, self-addressed return envelope (or e-mail return address), abstracts of less than 200 words (any format) should be submitted to Cathey Falvo, MD, MPH, Graduate School of Health Sciences, New York Medical College, Valhalla, NY 10595; tel (914) 993-4323; fax (914) 993-4434; falvo@nymc.edu (internet).

Abstracts must be *received* by September 12, 1994. Decisions will be made by September 19, 1994.

Self-Rated Health and Survival: A 7-Year Follow-Up Study of Australian Elderly

John McCallum, DPhil, MPhil, Bruce Shadbolt, PhD, and Dong Wang, BSc

ABSTRACT

Objectives. This study tested the hypothesis, from North American findings, that global self-ratings of health predict survival for older Australians.

Methods. A stratified sample of Australians 60 years of age and older surveyed in 1981 was resurveyed in 1988. Cox proportional hazard general linear models were constructed separately for men and women to predict survival over 7 years.

Results. Better self-ratings of health had an incremental association with survival for women, but only men with poor ratings had significantly worse survival than others. After major illnesses, comorbidities, disability, depression, and social support were controlled for, poor ratings of health for both men and women were not significantly different from excellent ratings in predicting survival. Only women's good and fair health ratings remained significant predictors.

Conclusions. People rate their health as poor on the objective basis of illness and disability. Australian findings show gender differences relative to North American results; methodological differences and site and gender variability in health profiles are discussed as reasons for the varying results. (*Am J Public Health.* 1994;84:1100–1105)

Introduction

Global self-ratings of health are simple survey instruments, but evidence is growing that they are powerful, independent predictors of survival among the aged. In an article published in the *American Journal of Public Health* in 1982, the researchers of the Manitoba Longitudinal Study on Aging[1] first reported that self-ratings of health predicted survival independent of objective health status indicators. Self-ratings of health had been used as an economical substitute for clinical assessments,[2,3] but the Manitoba result showed that self-rated health predicted survival in its own right. Sociological claims that even ungrounded beliefs have real effects[4] and that health beliefs determine health outcomes[5] were given credence by the result.

An editorial[6] in the same issue raised concerns about the effectiveness of controls for health status. Subsequent US studies using data from different sites (Alameda County in California[7]; New Haven, Conn[8,9]; and Iowa and Washington counties in Iowa[8]), along with the first National Health and Nutrition Examination Survey (NHANES I),[10] the Longitudinal Study on Aging, and a study conducted in Kiryat One, Israel,[12] further tested the relationship between self-ratings of health and survival. Increasingly complex models with ever richer measures of health and other predictors of survival found, with the exception of NHANES I, that self-rated health remained a significant, independent predictor of survival.

Theoretical explanations followed rather than led the investigations. It was proposed that, while older people take account of medical information in self-assessments, health experts do not take account of psychosocial factors or subjective well-being in their assessments.[13] When a wide range of psychosocial and well-being measures were included in the new studies, the self-rated health effect remained significant, indicating that the relationship between self-rated health and survival is independent of psychosocial measures whether or not they enter into physicians' diagnoses.

Three specific explanations have been proposed.[14] First, negative perceptions can stimulate the release of chemicals that compromise the immune system (the psychoneuroimmunological explanation). Second, poor self-perceptions may derive from self-detection of health conditions in preclinical phases. Third, these perceptions may be associated with delay or resistance in taking actions to protect and maintain one's health.

These explanations do not find consistent support in existing results. First, the lack of significance of depression in various models is inconsistent with a psychoneuroimmunological explanation. Second, in NHANES I,[10] self-ratings of health predicted early and late mortality equally well; thus, it was unlikely that personal knowledge of undiagnosed conditions explained the mortality prediction. Third, studies with extensive risk factor behavior controls[8,9] also showed that the effect is independent of risky health behaviors. Consequently, Idler and Kasl concluded that health perceptions have direct and independent effects on survival but that these effects work through un-

The authors are with the National Centre for Epidemiology and Population Health, The Australian National University, Canberra.

Requests for reprints should be sent to John McCallum, DPhil, MPhil, National Centre for Epidemiology and Population Health, The Australian National University, Canberra ACT 2601, Australia.

This paper was accepted October 4, 1993.

known processes: "Health perceptions may represent a unique construct, and a unique process, not reducible to other already known constructs and processes."[8]

While existing discussions have stressed the consistency of the observed relationship between self-ratings of health and survival, results by site and gender have been inconsistent across studies in some details. The NHANES I analysis reported that self-rated health was associated with survival over a 12-year period among middle-aged men but not among elderly men or women. Possible reasons offered for the lack of results for the 65- to 74-year age group were the more extensive list of physician diagnoses and risk factor measures and the use of a heterogeneous national sample. In Iowa,[9] only the extreme poor/bad categories remained significant for women, and only the poor/bad and fair categories remained significant for men. Only two of the seven investigations mentioned earlier, the Alameda County study[7] and the Longitudinal Study on Aging,[11] showed larger adjusted odds ratios for women than men for the worst self-rated health category included. The Kiryat One results[12] from Israel matched those from most US studies with significant adjusted odds ratios that were larger for men than for women.

Variability by site and gender is consistent with the fact that self-assessments have no absolute reference point. People use personal, peer group, and objective medical information to make a global assessment. It follows that self-ratings of health can be based on different information for men and women at different phases of the life cycle and in different social contexts. This raises the question addressed here: Will self-rated health predict survival in a similar way among elderly individuals residing in Australia?

Methods and Data

Hazard of death was analyzed as a response variable to baseline self-assessments of health. After examination of how well self-rated health predicted survival, adjustments were made for various confounding factors. Survival analyses were conducted in LIMDEP[15]; regression was used as a partial likelihood procedure to model risk. On the basis of Cox[16] proportional hazards, inferences can be made about the risk multiplier $g(x)$ without involving the function $l(t)$, which provides the underlying shape of the variation of hazard with time. The hazard of death at time t, denoted by $h(t, x)$, is dependent on covariables denoted by x, as follows:

$$h(t, x) = l(t)g(x).$$

The analysis used data from the Ageing and the Family Project Survey.[17] This survey, conducted in November 1981, involved interviews with a random sample of 1050 people living in Sydney, Australia, who were 60 years of age and older. After enumeration of households in randomly selected districts, eligible respondents were stratified by age and living arrangements. The sample strata were designed to oversample very old people not living with other older people: 25% of persons 60 through 64 years of age were sampled, along of with 50% of those 65 through 74 years of age, 50% of those 75 years of age or older if they lived with someone else 60 years of age or older, and 100% of those 75 years of age or older if they did not live with someone 60 years of age or older.

In 1988,[18] a follow-up questionnaire was administered to the survivors from 1981. People who died over the 7-year period (22% of respondents) were identified from state death records and their dates of death recorded. Seventy percent of the respondents participated in the 1981 baseline survey, 87% of whom agreed to follow-up. The status of 91% of the follow-up respondents was determined, and the 9% who were untraceable were excluded from analyses. Analysis of dropouts indicated no significant biases by age, sex, or other major demographic characteristics.

Measures

Self-rated health was measured by the question "Would you say your overall health is excellent, good, fair, or poor?" Three dummy variables were used to compare the latter three states with "excellent."

Controls for age (5-year groups) and living with someone 60 years of age or older were necessary in dealing with the sample strata. The few interviews conducted by proxy because of frailty or language difficulty were controlled by use of a dummy variable.

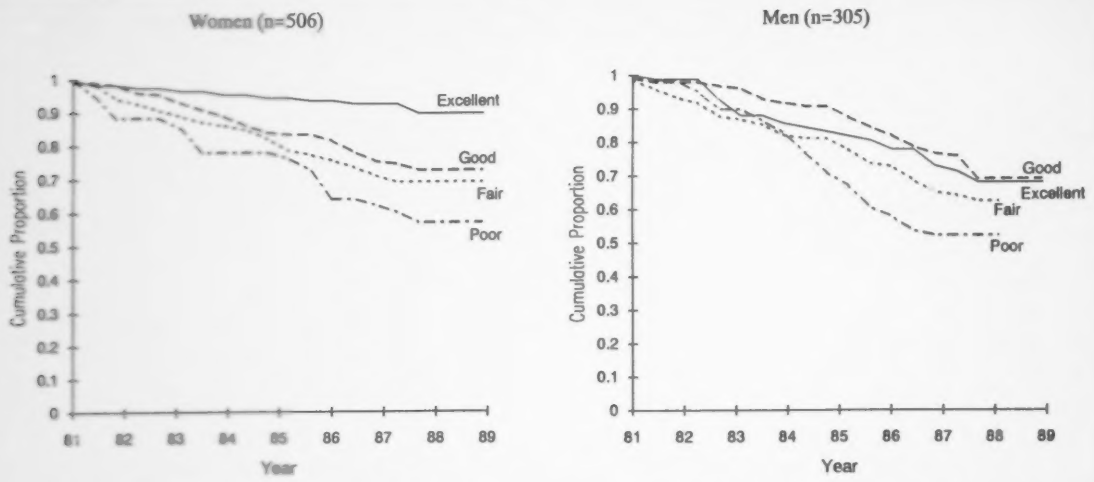
Various other measures were included in models to examine confounding effects on self-rated health and survival. First, the marital status categories widowed, never married, and divorced/separated were compared with married status. Possible confounding survival advantages reported for married people may be a consequence of either "health selection" of people into marriage or "social causation" during marriage.[19]

Older people born in Greece and Italy were identified by a dummy variable. About 35% of the Australians 65 years of age and older were born overseas, including some who were members of immigrant groups with survival advantages. The most notable group was that composed of Greek- and Italian-born immigrants, who had a lower risk of coronary heart disease and some cancers.[20] Other immigrant groups had very small numbers or no survival advantage.

Ownership of a private dwelling, a general indicator of wealth, relief from rent, and income, was used here as a single variable to control for possible confounding effects of class and income. Social class measures have been shown in Australia to have an inverse association with survival,[21] but this class effect disappears with age.[22] Income-based measures of class among the elderly are affected by the flat-rate Australian retirement income system and provide poor indications of lifelong advantages.

To control for objective health status, respondents were asked "Have you had a serious illness or accident in the past 5 years? What was that?" Survey respondents generally provide accurate reports of their health status, except when a disease is diagnostically complex.[23-25] Sufficient numbers of our respondents reported circulatory, respiratory, and musculoskeletal conditions; neoplasms; and falls. The remaining diseases reported were classified as "other," and the null category for comparison was the absence of a condition in the last 5 years. People who reported more than one serious illness in the last 5 years were defined as having a serious comorbidity unless one of the illnesses was musculoskeletal. Whether someone had been ill in bed at home in the last 6 months was also used as a dummy variable for minor and undiagnosed illness.

Disability was measured by a mix of items focusing on activities of daily living and instrumental activities of daily living, including bathing, doing housework, climbing stairs, going out of the house, using public transportation, and being able to walk half a mile. A composite measure of functional health was constructed by



Note. The proportions surviving have been adjusted for variations in the age structure of the samples by means of direct standardization, in which the age distribution of women exposed at the baseline survey who rated their health as excellent is the standard.

FIGURE 1—Age-standardized cumulative proportions of survival over the study period within each self-rated health group (Excellent, Good, Fair, and Poor).

adding one point for each item performed without difficulty.

Depression was measured with the Philadelphia Geriatric Center Morale Scale.[26] Because a connection has been observed between depression and survival, depression was a potential confounding factor; however, there is controversy about its causal direction.[27-29] Depressed (2 to 8 symptoms) and very depressed (9 to 12 symptoms) respondents were compared with those with no symptoms or 1 symptom. The missing values on the 12-item morale scale were modeled as a separate category.

The effects of social support on mortality are well established,[30,31] and these effects were other potential confounders in this study. Three indicators of emotional support were used: having someone to confide in, having someone who makes you feel needed, and having someone with whom you share common interests. Since all three measures had high positive intercorrelations, a composite emotional support measure was constructed by adding a point for each positive response. Social participation was indicated by four items: having someone come to visit, going out to visit others, going on outings and social gatherings with others, and attending clubs and other social organizations. A composite measure of social participation was constructed by assigning each positive answer a score of 1.

Results

Overall, 76% of women, in comparison with 67% of men, survived the follow-up period. Figure 1 shows that, after standardizing for age, the relationship between self-rated health and survival was relatively well ordered for women; for men, however, only the poor rating was clearly different from others. Men, especially those with excellent self-rated health, showed a lower probability of survival than did women. Men with good ratings had survival probabilities similar to those of men with excellent ratings.

The results in Figure 1 justified further analysis of the effects of self-rated health after controlling progressively for a broad range of factors that predict survival. For women (see Table 1), the addition of sample characteristics and demographics (model 1) did not significantly alter the relationship between self-rated health and survival; mortality risks increased incrementally with poorer self-ratings of health. The relative risks (exponentiation of the regression parameters) were 3.2 (95% confidence interval [CI] = 1.6, 6.5) for good vs excellent, 3.8 (95% CI = 1.8, 7.9) for fair vs excellent, and 5.3 (95% CI = 2.4, 11.7) for poor vs excellent. After controlling for serious illnesses and accidents (model 2), self-rated health remained a significant predictor, but the relative risk associated with poor self-rated health was reduced to 3.2 (95% CI = 1.3, 7.4). Controlling for disability (model 3) further reduced the effect for poor self-rated health to the point at which it was no longer significant at the .05 level. Depression did not add significantly to model 4, but emotional support was significant. The relative risks of death for women in model 4 were 3.2 (95% CI = 1.5, 6.7) for good vs excellent, 2.4 (95% CI = 1.1, 5.2) for fair vs excellent, and 1.9 (95% CI = 0.7, 5.1) for poor vs excellent.

For men (see Table 2), after controlling for sample characteristics and demographics (model 1), only the coefficient for poor self-rated health was significantly different from that for excellent self-rated health. Men's relative risks were 0.8 (95% CI = 0.4, 1.4) for good vs excellent, 1.5 (95% CI = 0.8, 2.7) for fair vs excellent, and 3.1 (95% CI = 1.7, 5.8) for poor vs excellent. The result for the good self-rated health category indicates that men with good ratings had a lower (but nonsignificant) risk than the baseline

TABLE 1—Parameter Estimates and Standard Errors for Women in Four Cox Proportional Hazard Regression Models: 7-Year Follow-Up of the Ageing and the Family Project Survey

Factor	Coefficient (SE)			
	Model 1: Baseline	Model 2: Illness	Model 3: Disability	Model 4: Full
Controls				
Age (5-y groups)	.557 (.083)***	.511 (.085)***	.428 (.089)***	.408 (.089)***
Live with person 60+ years of age	−.107 (.249)	−.041 (.259)	.015 (.257)	−.001 (.257)
Proxy interview	.140 (.222)	.037 (.231)	.057 (.231)	−.048 (.238)
Demographics				
Greek/Italian born[a]	. . .	. . .	. . .	. . .
Home owner	.058 (.207)	.073 (.212)	.112 (.212)	.225 (.216)
Widow vs married	−.078 (.264)	−.169 (.268)	−.149 (.268)	−.180 (.269)
Divorced/separated vs married	−.159 (.564)	−.688 (.612)	−.887 (.617)	−.757 (.603)
Never married vs married	.340 (.389)	.266 (.400)	.270 (.398)	.151 (.403)
Self-rated health				
Poor vs excellent	1.664 (.406)**	1.148 (.437)**	.693 (.463)	.665 (.496)
Fair vs excellent	1.330 (.374)***	1.126 (.386)**	.918 (.392)*	.863 (.404)*
Good vs excellent	1.164 (.364)***	1.160 (.370)**	1.087 (.371)**	1.154 (.380)**
Serious illnesses (in last 5 y)				
Circulatory diseases		.742 (.291)*	.752 (.298)*	.628 (.318)*
Neoplasms		1.116 (.435)*	1.189 (.424)**	1.248 (.421)**
Respiratory diseases		.879 (.403)*	.710 (.405)	.709 (.396)
Falls		.711 (.378)	.507 (.388)	.600 (.401)
Musculoskeletal disease		−.108 (.387)	−.134 (.387)	−.077 (.383)
Other serious disease		.185 (.306)	.152 (.301)	.114 (.303)
Other illness				
Comorbidities		.389 (.392)	.345 (.388)	.371 (.392)
Ill in bed at home (in last 6 mo)		.070 (.248)	.039 (.247)	−.025 (.251)
Function: disability score (current)			−.152 (.052)**	−.120 (.056)*
Depression				
Normal				.289 (.317)
Depressed				.162 (.544)
Refused				.330 (.437)
Social support				
Emotional support				−.243 (.121)*
Social participation				−.111 (.098)
Log likelihood (*df*)[b]	−669.27 (10)	−655.16 (18)	−650.99 (19)	−644.57 (24)
Log likelihood excluding self-rated health (*df*)[c]	−680.74 (7)	−662.03 (15)	−656.69 (16)	−650.62 (21)

Note. n = 506.
[a]Numbers too small for inclusion.
[b]Log likelihood (slope = 0) = −716.72.
[c]The same model less the three self-rated health dummies.
*$P < .05$; **$P < .01$; ***$P < .001$.

hazard for men with excellent ratings. After controlling for serious illnesses and accidents (model 2), self-rated health remained significant, but the relative risk for poor self-rated health was reduced to 2.6 (95% CI = 1.2, 5.5). Controlling for disability (model 3) further reduced the relative risk for poor self-rated health to the point at which it was no longer significant at the .05 level. Depression and social support did not significantly add to model 4. The relative risks for men in model 4 were 0.8 (95% CI = 0.4, 1.5) for good vs excellent, 1.1 (95% CI = 0.6, 2.2) for fair vs excellent, and 1.7 (95% CI = 0.7, 4.3) for poor vs excellent.

The addition of the blocks of factors to the models (shown in Tables 1 and 2) made a significant contribution to log likelihood differences in each case, except for the additions to models 3 and 4 for men. The log likelihood values without self-rated health (calculated for each model by excluding self-rated health dummies) further show that self-rated health dummies added significantly to all models except models 3 and 4 for men.

Discussion

The results confirm that self-rated health is related to survival for older Australians. The finding that the survival effects of poor ratings of health were explained mostly by serious illnesses and disability is consistent with the model of health status recently specified by Johnson and Wolinsky.[32] They reported evidence for a natural progression of effects from body to mind as diseases take their toll on activities of daily living and dampen the perception of good health. Thus, the effects of poor ratings observed here fit a medical model of survival.

There was, however, variation in the effects of intermediate ratings between Australian men and women. After the addition of controls, men's ratings did not

TABLE 2—Parameter Estimates and Standard Errors for Men in Four Cox Proportional Hazard Regression Models: 7-Year Follow-Up of the Ageing and the Family Project Survey

Factor	Coefficient (SE)			
	Model 1: Baseline	Model 2: Illness	Model 3: Disability	Model 4: Full
Controls				
Age (5-y groups)	.456 (.093)***	.470 (.098)***	.456 (.099)***	.462 (.102)***
Live with person 60+ years of age	.043 (.342)	.198 (.345)	.096 (.355)	.092 (.361)
Proxy interview	.374 (.125)	.133 (.259)	.047 (.268)	.053 (.306)
Demographics				
Greek/Italian born	−1.984 (.1021)*	−2.101 (1.045)*	−2.202 (1.051)*	−2.246 (1.059)*
Home owner	−.148 (.241)	.005 (.258)	.041 (.259)	.064 (.264)
Widow vs married	.279 (.314)	.343 (.311)	.432 (.318)	.458 (.320)
Divorced/separated vs married	.319 (.527)	.535 (.540)	.596 (.543)	.512 (.553)
Never married vs married	.421 (.431)	.515 (.459)	.444 (.464)	.540 (.507)
Self-rated health				
Poor vs excellent	1.129 (.318)***	.953 (.379)*	.710 (.426)	.563 (.456)
Fair vs excellent	.375 (.316)	.244 (.335)	.198 (.337)	.106 (.352)
Good vs excellent	−.214 (.305)	−.186 (.317)	−.228 (.320)	−.252 (.323)
Serious illnesses (in last 5 y)				
Circulatory diseases		.278 (.347)	.265 (.347)	.270 (.349)
Neoplasms		1.146 (.573)*	1.162 (.577)*	1.250 (.614)*
Respiratory diseases		.467 (.548)	.204 (.587)	.245 (.601)
Falls		1.628 (.578)**	1.618 (.578)**	1.691 (.588)**
Musculoskeletal disease		.071 (.447)	−.028 (.453)	−.086 (.467)
Other serious disease		−.046 (.343)	−.038 (.343)	−.009 (.349)
Other illness				
Comorbidities		.706 (.470)	.729 (.471)	.773 (.482)
Ill in bed at home (in last 6 mo)		.284 (.293)	.267 (.298)	.263 (.301)
Function: disability score (current)			−.107 (.077)	−.103 (.077)
Depression				
Normal				.144 (.294)
Depressed				.482 (.472)
Refused				.211 (.491)
Social support				
Emotional support				.053 (.131)
Social participation				−.084 (.122)
Log likelihood (*df*)[a]	−489.48 (11)	−480.09 (19)	−479.13 (20)	−478.36 (25)
Log likelihood excluding self-rated health (*df*)[b]	−502.50 (8)	−485.26 (16)	−482.01 (17)	−480.20 (22)

Note. n = 305.
[a]Log likelihood (slope = 0) = −519.37.
[b]The same model less the three self-rated health dummies.
*$P < .05$; **$P < .01$; ***$P < .001$.

differentiate the risks of survival. The persistence of effects on survival of women's good and fair ratings, after the fitting of controls, adds a degree of complexity to the model. Australian women appear to use information that is not used by men for the intermediate ratings. The addition of healthy life-style indicators to the model (not available here) may reduce the effects of these intermediate categories for women.

These findings add to the existing evidence of variability in North American studies. For example, no effects after fitting controls were observed for the 65- to 74-year age group in the NHANES I analysis,[10] and, in Iowa,[9] the poor/bad and fair ratings were significant for men, but only the extreme poor/bad rating was significant for women. Furthermore, in this study relative risks were higher for women than for men, a finding similar to that of only two existing studies.[7,11]

How are the different findings between studies to be explained? Three possibilities are different methods, different health profiles for populations, and different health perceptions by sex and sites. The study methods differ in specific questions used and response categories, length of follow-up, statistical methods, and sampling strategies. Nonetheless, studies indicate general consistency in the relationship between self-rated health and survival. However, the Longitudinal Study on Aging authors[11] noted that men's results are best behaved in studies that use the extreme category "bad" rather than the poor category. Consistency in methods is a desirable goal for future studies of the relationship between self-ratings of health and survival.

As a result of different sex role socialization and cultures, differences may exist between sites in the ways that people combine different types of disease and disability information to make self-assessments. The concentration in existing studies on simple global assessments needs to be supplemented by testing of detailed models of the process of self-

assessment such as the model proposed by Johnson and Wolinsky.[32] On the basis of the results of this study, it would be expected that such processes differ systematically by gender and site.

Furthermore, self-rated health can embody actual differences in health information (e.g., men and women have different "objective" health profiles, as do people in different areas). National data for older Australians in 1981 and 1988 showed that death rates for coronary heart disease were two to three times higher among men than among women.[33] The dominant risk factors for coronary heart disease, such as smoking and hypertension, have large effects for men. For women, for whom single risk factors are not dominant, more variables with smaller coefficients reach significance in the model.[34] The prevalent, acute nature of heart attacks for this cohort of men may create the basis for less gradation in men's self-ratings with respect to survival relative to women and the better match of Australian men's perceptions of poor health to a "medical model."

The persistent effect of self-rated health among the elderly despite extensive health controls may be due to the inability of medical and social science to adequately model the complex, multiple, interacting illnesses of elderly people. It is known that risk analysis does not account for frailty (the problem of many variables each contributing individually a small amount to total variability but together making a significant difference).[35] We await statistical developments in the measurement of frailty to test this possibility.

Also, there are no economical survey measures of system aging or senescence (system failure distinct from failure of body parts and specific illness) independent of chronological age. Hence, senescence is generally omitted from health status models. The possibility that self-perceptions indicate frailty or senescence suggests that it is not yet time to claim that the self-rated health effect is unique and the process unknown.[8] Senescence may well be better recognized by aging people than by any existing survey measures.

The two interests in self-ratings of health are confirmed by Australian evidence for the elderly. First, self-rated health is an economical indicator of health defined by severe illnesses and disability. Second, it adds finer-graded information about health related to survival, but only for women. The specific variability observed in Australian and other findings points to the need for more research to explore the structure of health status among older adults. □

Acknowledgments

The 1981 Ageing and the Family Project Survey was conducted by the Research School of Social Sciences of The Australian National University with supplemental funding from the Commonwealth Department of Health. The 1988 follow-up was conducted by the National Centre for Epidemiology and Population Health of The Australian National University.

References

1. Mossey JM, Shapiro E. Self-rated health: a predictor of mortality among the elderly. *Am J Public Health.* 1982;72:800–808.
2. Fillenbaum GG. Social context and self-assessments of health among the elderly. *J Health Soc Behav.* 1973;20:45–51.
3. LaRue A, Bank A, Jarvik L, et al. Health in old age: how do physicians' ratings and self-ratings compare? *J Gerontol.* 1979;34: 687–691.
4. Thomas WI, Znaniecki F. *The Polish Peasant in Europe and America.* New York, NY: Dover; 1958.
5. Eisdorfer C, Wilkie F. Stress, disease, aging, and behavior. In: Birren J, Warner Schaie K, eds. *Handbook of the Psychology of Aging.* New York, NY: Van Nostrand Reinhold; 1977.
6. Muller C. Health status and survival needs of the elderly. *Am J Public Health.* 1982;72: 789–792.
7. Kaplan GA, Camacho T. Perceived health and mortality: a nine year follow-up of the human population laboratory cohort. *Am J Epidemiol.* 1990;117:292–304.
8. Idler EL, Kasl S. Health perceptions and survival: do global evaluations of health status really predict mortality? *J Gerontol.* 1991;46:S55–S65.
9. Idler EL, Kasl SV, Lemke JH. Self-evaluated health and mortality among the elderly in New Haven, Connecticut, and Iowa and Washington counties, Iowa, 1982–1986. *Am J Epidemiol.* 1990;82:91–103.
10. Idler EL, Angel RJ. Self-rated health and mortality in the NHANES-1 epidemiologic follow-up study. *Am J Public Health.* 1990; 80:446–452.
11. Wolinsky FD, Johnson RJ. Perceived health status and mortality among older men and women. *J Gerontol.* 1992;47:S304–S312.
12. Kaplan G, Barell V, Lusky A. Subjective state of health and survival in elderly adults. *J Gerontol.* 1988;43:S114–S120.
13. Sagan LA. *The Health of Nations. True Causes of Sickness and Well-Being.* New York, NY: Basic Books; 1987:174.
14. Rakowski W, Mor V, Hiris J. The association of self-rated health with two year mortality in a sample of well elderly. *J Aging Health.* 1991;3:527–545.
15. Greene WH. *LIMDEP Users Manual.* New York, NY: New York University; 1990.
16. Cox DR. Regression models and life tables. *J R Stat Soc Ser B.* 1972;34(pt 2):187–202.
17. Kendig HL, ed. *Ageing and Families. A Support Networks Perspective.* Sydney, Australia: Allen and Unwin; 1986.
18. McCallum J, Kendig HL, Freeman E, et al. *Ageing and Families 7 Years After: Data from a Seven Year Follow-Up of Older Australians.* Canberra, Australia: National Centre for Epidemiology and Population Health; 1991.
19. Wyke S, Ford G. Competing explanations for associations between marital status and health. *Soc Sci Med.* 1992;34:523–532.
20. Powles J, Gifford S. How healthy are Australia's immigrants? In: Reid J, Trompf P, eds. *The Health of Immigrant Australia.* Sydney, Australia: Harcourt Brace Jovanovich; 1991.
21. Moss JR, McMichael AJ. Inequalities in health. The epidemiological evidence—and the gaps. In: McNeil J, King R, Jennings G, Powles J, eds. *A Textbook of Preventative Medicine.* Melbourne, Australia: Edward Arnold; 1991.
22. Australian Institute of Health. *Australia's Health 1992.* Canberra, Australia: Australian Government Publishing Service; 1992.
23. Colditz GA, Martin P, Stampfer MJ, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol.* 1986;123:894–900.
24. Harlow SD, Linet MS. Agreement between questionnaire data and medical records. *Am J Epidemiol.* 1989;129:233–248.
25. Coulter A, McPherson K, Elliott S, Whiting B. Accuracy of recall of surgical histories: a comparison of postal survey data and general practice records. *Community Med.* 1985;7:186–189.
26. Lawton MP. The Philadelphia Geriatric Center Morale Scale: a revision. *J Gerontol.* 1975;30:85–89.
27. Jorm AF, Henderson AS, Kay DW, Jacomb PA. Mortality in relation to dementia, depression and social integration in an elderly community sample. *Int J Geriatr Psychol.* 1991;6:5–11.
28. Murphy JM, Monson RR, Olivier DC, et al. Affective disorders and mortality. A general population study. *Arch Gen Psychiatry.* 1987;44:473–480.
29. Murphy E, Smith R, Lindesay J, Slattery J. Increased mortality rates in late life depression. *Br J Psychiatry.* 1988;152:347–353.
30. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *Am J Epidemiol.* 1979;109:186–204.
31. House JS, Landis KR, Umbertson D. Social relationships and health. *Science.* 1989;241:540–544.
32. Johnson RJ, Wolinsky FD. The structure of health status among older adults: disease, disability, functional limitation, and perceived health. *J Health Soc Behav.* 1993;34:105–121.
33. McCallum J. Health: the quality of survival in old age. In: Australian Institute of Health, *Australia's Health 1990.* Canberra, Australia: Australian Government Publishing Service; 1991.
34. Simons L, Friedlander Y, McCallum J, et al. The Dubbo Study of the Health of the Elderly: correlates of coronary heart disease at study entry. *J Am Geriatr Soc.* 1991;39:584–590.
35. Hougaard P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika.* 1986;71: 75–83.

Work-Related Traumatic Brain Injury in Washington State, 1988 through 1990

ABSTRACT

Objectives. Brain trauma is a major cause of disability in the United States, especially among young adults. Work-related brain trauma cases represent a subpopulation that may be amenable to intervention efforts, but largely because of unavailability of data, this group has not previously been studied.

Methods. Washington State workers' compensation data were used to identify brain trauma cases and to describe incidence rates with respect to age, gender, and industrial classification. Cause of injury was evaluated for industrial classifications with numerous cases.

Results. Washington State workers experienced 301 brain traumas over a 3-year period (9.4/100 000 full-time-equivalent workers per year). One third of all traumatic brain injury claims were concentrated among seven industrial classifications. Cause of injury was highly correlated with specific industrial classifications.

Conclusions. This study demonstrates the usefulness of diagnosis-related group codes to identify cases of traumatic brain injury. These injuries are concentrated among a few industries, with a dominant specific cause for each industry. This suggests that intervention methods may be efficiently introduced to reduce traumatic brain injuries in the workplace. (*Am J Public Health.* 1994;84:1106–1109)

Nicholas J. Heyer, PhD, and Gary M. Franklin, MD, MPH

Introduction

Brain trauma is a major cause of disability, especially among young adults. The costs of brain trauma in the United States were estimated at $4 billion in 1980.[1] Although many studies have generated traumatic head and/or brain injury incidence rates for general populations,[2] none have focused on persons with work-related brain trauma as a subpopulation potentially amenable to prevention and control efforts.

The dearth of information on work-related brain trauma is most likely related to the difficulty in identification of these cases. Occupation codes associated with hospital discharge data and death certificates are typically underreported or unreliable. One US study[3] reported that 4.4% of all brain traumas (7.3% of those rated as severe) were work related. In this population-based study, Annegers et al.[3] used the unique records-linkage system of the Mayo Clinic in Rochester, Minn, to identify cases; however, skull fractures without brain injury were included, and criteria for work-relatedness were not described. A Norwegian study using similar case definitions reported that 4.8% of incident brain trauma cases were work related.[4]

We used an administrative database from the Washington State workers' compensation system to determine: (1) the most useful codes for ascertainment of hospitalized brain trauma cases (e.g., diagnosis-related group [DRG] codes; *International Classification of Diseases,* Ninth Revision [ICD-9], codes; or American National Standards Institute [Z16.2] codes); (2) age-, gender-, and industry-specific incidence rates of work-related hospitalized traumatic brain injury; and (3) whether industry-specific variability appears substantial enough for surveillance and targeted prevention efforts.

Methods

The Workers' Compensation Data System

The Department of Labor and Industries is the sole regulator of workers' compensation coverage for non-federally covered workers in Washington State. Approximately two thirds of these workers are covered through the State Fund; the remainder are covered through approximately 365 large companies that are able to self-insure. This study uses the extensive claim and medical bill payment information maintained by the Department of Labor and Industries on State Fund-covered workers. Data on workers covered by self-insured companies are inadequate for research purposes and are not included in this study.

The following selected variables from the Department of Labor and Industries database were reviewed: (1) Department claim data: age, gender, Z16.2 nature of injury codes, and industry codes (Washington Industrial Classification)[5]; and (2) Medical bill payment data: hospital ICD diagnostic codes, hospital DRG codes, and length of hospital stay. The DRG codes used are from Version 8 of the Medicare DRG grouper.

Nicholas J. Heyer and Gary M. Franklin are with the Department of Environmental Health, University of Washington School of Public Health and Community Medicine, Seattle, Wash. Gary M. Franklin is also with the University of Washington School of Medicine (Neurology), Seattle, and the Washington State Department of Labor and Industries, Olympia.

Requests for reprints should be sent to Gary M. Franklin, MD, MPH, Department of Environmental Health, University of Washington School of Public Health and Community Medicine, 1914 N 34th St, Suite 101, Seattle, WA 98103.

This paper was accepted October 21, 1993.

TABLE 1—Medical Record Review of Probable Cases, by Selection Factor Status: Traumatic Brain Injury, Washington State Workers' Compensation, 1988 through 1990

Selection Factor	No. Included (Excluded)[a] Claims
Definite DRG	294 (5)
DRG only	35 (3)
DRG + Z16.2 only	5 (1)
DRG + ICD-9 only	131 (1)
DRG + ICD-9 + Z16.2	123 (0)
Definite ICD-9 + Z16.2	7 (1)
Total	301 (6)

Note. DRG = diagnosis-related group code; ICD-9 = *International Classification of Diseases*, Ninth Revision, code.
[a]These additional claims were excluded for jurisdictional reasons.

TABLE 2—Case Distribution, by Calendar Year of Hospitalized Brain Injury: Washington State Workers' Compensation, 1988 through 1990

Year	Work Hours	Total FTE	Cases, n	Rate[a]	Average LOS[b]	One Day Only, %[c]
1988	2 009 431 816	1 004 715.90	104	10.35	6.95	30
1989	2 133 765 047	1 066 882.52	114	10.69	7.10	21
1990	2 254 569 192	1 127 284.59	83	7.36	6.01	2
Total	6 397 766 055	3 198 883.01	301	9.41	6.75	20

Note. FTE = full-time equivalent; LOS = length of stay.
[a]Incidence rate per 10^3 FTE/year.
[b]Average length of stay in days—index hospitalization.
[c]Percentage with index hospitalization of only 1 day.

Case Definition

A case of traumatic brain injury was considered definite if any of the following were documented in the medical records in association with a head trauma: clinical or imaging evidence of brain injury, a medical diagnosis of concussion, or a clear description of loss and/or alteration of consciousness. A diagnosis of closed head injury or fracture with none of the above documented findings was not considered sufficient for inclusion.

Case Ascertainment

All State Fund-insured claims requiring hospitalization were identified for calendar years 1988 through 1990. Hospital DRG codes, hospital ICD codes (up to five codes), and Z16.2 codes for part of body injured were classified a priori to indicate possible traumatic brain injury cases. These codes were checked both singly and in combination for their specificity by a review of medical records from a representative sample of 40 claims.

Based on the review of claims, probable brain trauma cases were defined as those with a DRG specifying brain trauma and/or those with both ICD and Z16.2 codes specifying brain trauma. Final selection of claims for this study was based on the following criteria: (1) a review of the medical records for each probable claim identified by the case definition described above; (2) date of injury between January 1, 1988, and December 31, 1990; and (3) hospitalization within 7 days of the date of injury.

Age-, Gender-, and Industry-Specific Rates

Hours worked per year are reported for individual industry groupings (Washington Industrial Classification). Full-time-equivalent workers were calculated for each industrial class by dividing the total hours reported by 2000.[5] Hours reported by industry class are not reported by age and gender. Thus, denominators for age- and gender-specific rates were estimated from the 1990 census (Washington State Office of Financial Management). The proportion of the work force by age and gender was applied to the total number of hours reported to the Department of Labor and Industries.

Statistical Analysis

All rates were calculated as the number of incident brain trauma cases/100 000 full-time equivalents per year. Industry-specific rates were compared with the industrywide rate by using a rate ratio with a 95% confidence interval.[6]

Results

A total of 310 probable brain trauma cases met the date of injury and hospitalization within 7 days requirements of this study. A detailed review of these cases found that 301 qualified for this study. Table 1 shows the distribution of the 301 qualified cases as well as the 6 that were excluded for not meeting medical criteria, according to our DRG, ICD-9, and Z16.2 selection factors. All but 7 of the qualified cases would have been identified by DRG alone.

Based on 301 cases, the industrywide average annual incidence of work-related hospitalized traumatic brain injury was 9.4 cases/100 000 full-time equivalents (Table 2). The decline in hospitalized cases in 1990 is explained by a dramatic fall in the proportion of cases hospitalized for only 1 day.

Figure 2 shows the age-specific incidence for brain trauma injuries. The mean age of traumatic brain injury was 36 years (range, 14–75 years), with peaks of injury occurring in the 16–34-year and 55–64-year age groups. The incidence rate of brain trauma was 6.3 times higher for men than for women, the rate for men being 15.1 per 100 000 full-time equivalents and the rate for women being 2.4.

The majority of brain trauma cases resulted from falls (48.5%), from being struck by an object (26%), or from motor vehicle accidents (18.3%). The average length of the index hospitalization, a marker of case severity, averaged 6.8 days with little variation by external cause of injury (Table 3). Case severity appeared to be more clearly related to specific DRG codes (Table 4).

The seven industrial classifications with the highest number of traumatic brain injuries (six or more) accounted for one third (101/301) of all such cases (Table 5). Brain trauma cases tended to cluster by external cause of injury within industrial classes. For example, 85% (29/34) of cases in logging were due to being struck by an object, whereas 84% (21/25) of cases in wood-frame building construction, 86% (6/7) of cases in roof work, and 77% (10/13) of cases in orchard and dairy farms were due to falls from height. Also, 58% (11/19) of cases in trucking were associated with motor vehicle accidents. Industries with significantly elevated relative rates (RRs) include logging (RR = 36.7), roof work (RR = 17.1), garbage collection (RR = 13.5), street and road construction (RR = 12.4), trucking (RR = 10.0), dairy farms (RR = 8.9), interior finish carpentry (RR = 8.8), and building

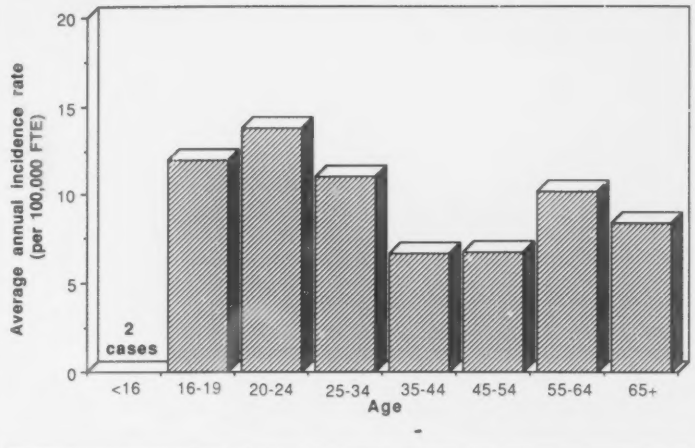


FIGURE 1—Occupational morbid brain injury rate, by age, Washington State, 1988 through 1990.

TABLE 3—Case Severity Distribution, by External Cause of Injury: Traumatic Brain Injury, Washington State Workers' Compensation, 1988 through 1990 (n = 301)

External Cause of Injury	%	Length of Stay[a]
Fall from elevation	39.9	6
Struck by	26.3	8
Highway motor vehicle accident	18.3	6
Fall on same level	8.6	5
Struck against	3.6	3
Caught in/under/between	1.6	7
Explosion	0.7	7
Highway, non–motor vehicle accident	1.0	27
Total	100.0	

[a]Average length of stay in days—index hospitalization.

TABLE 4—Hospital Length of Stay, by Selected DRG Code: Traumatic Brain Injury, Washington State Workers' Compensation, 1988 through 1990

DRG Code	Description of Injury	No. Cases	Length of Stay[a]
002	Craniotomy, trauma	70	16
	Trauma, stupor		
027	>1 h	29	9
028	<1 h—complication	22	6
029	<1 h—no complication	52	4
030	<1 h—age of <18 y	4	5
	Concussion		
031	Complication	22	3
032	No complication	92	3
033	Age of <18 y	3	2

Note. DRG = diagnosis-related group.
[a]Average length of stay in days—index hospitalization.

construction (RR = 7.2). See Table 5 for a more complete listing of information.

Discussion

Using a workers' compensation administrative database, we identified 301 incident cases of work-related traumatic brain injury involving hospitalization for the years 1988 through 1990 in Washington State. The average annual incidence of hospitalized brain trauma was 9.4 cases/100 000 full-time equivalents. The marked excess of males (ratio of males to females = 6.3:1) and the secondary age peak in the 55–64-year age range differ from the distributions reported for the general population, where ratios of males to females are usually in the 2 to 3:1 range, and a secondary peak occurs after 65 years of age.[2] These differences are likely due to different distributions of occupational and nonoccupational risk factors, but may also be partly explained by differences in underlying population distributions.

The proportion of head injuries ascribed to an industrial cause has ranged from 4.4% to 4.8% in population-based studies[3,4] to 8% to 14% in studies from regional referral centers.[7–11] The age-adjusted annual incidence rate reported for work-related head trauma from Rochester, Minn, was 11/100 000, similar to the rate reported here specifically for brain trauma. In contrast to the Rochester study, the current investigation excluded cases of skull fracture that had no medical documentation of clinical brain injury.

Work-related brain trauma cases were most efficiently identified by hospital DRG codes. Both ICD and Z16.2 codes, alone and in combination, performed worse than DRG codes in terms of their unique contribution to case finding. DRG codes are related to prospective payment for hospitals and as such may be more accurately recorded. In addition, payment is strongly tied to expected duration of hospitalization for specific diagnostic categories. This precept was validated in this study; mean length of index hospitalization was strongly related to increasingly severe DRG categories. Patients with uncomplicated concussion averaged 3 days of hospitalization, whereas those with stupor lasting longer than 1 hour (but not requiring surgery) averaged 9 days in hospital.

As a prognostic marker of brain trauma severity, DRG codes may perform better than ICD codes. One prior study observed only 48% agreement between ICD codes and documented injury severity for head/neck injuries; ICD codes were far better in predicting severity of extremity injuries (74% agreement).[12] Further validation of the prognostic utility of DRG codes in predicting outcome of traumatic brain injury in population-based studies may be warranted. The most commonly used prognostic indicator of outcome, the Glasgow Coma Scale,[13] could not be reconstructed from the available medical record data in this study.

The mean length of stay for the index hospitalization reported here (6.8 days) is comparable to the length of stay reported in older studies[1,4] and may be a reasonable marker of injury severity.[14] The decline in rate of brain trauma admissions between 1988 and 1990 can be attributed entirely to a fall in 1-day hospital admissions. The proportion of such admissions fell from 30% in 1988 to only 2% in 1990. This decline is consistent with a sentinel effect related to recent, more restrictive Medicare billing policies for 1-day hospital admissions.

One third of all brain trauma cases occurred among only seven industrial classes. High relative risks were found for logging, roofing, garbage collecting, trucking, farming, and construction. External cause was related to industry type, so that falls were common in roofing, farming, and construction; motor vehicle accidents were common in trucking; and being struck by an object was the most common mode of injury in logging. Such clustering of traumatic brain injuries by external cause within industries and the large number of cases in relatively few industries provide an opportunity to target safety activities that would be a cost-effective method of preventing work-related traumatic brain injuries. Although work-related brain trauma most likely accounts for only 5% to 10% of all reported cases of traumatic brain injury, existing occupational safety and health resources could efficiently address this problem in the United States.

This is one of several recent studies undertaken to determine the usefulness of a workers' compensation database to generate population-based rates of work-related injuries.[5,15,16] Although there are some limitations to such studies (underreporting, misclassification), use of such extensive databases to clearly define variability among industries and thus to target further efforts at prevention and control is warranted. Our inability to access data on federal workers or on workers in the largest, self-insured companies would make extrapolation of the rates presented here to those venues difficult. The advantages of the Washington State system, excellent availability of diagnostic codes and of hours worked reported by industry class, have allowed an increased degree of specificity in studies using workers' compensation data. □

TABLE 5—Rates of Brain Trauma, by Industrial Class[a]: Traumatic Brain Injury, Washington State Workers' Compensation, 1988 through 1990

Industrial Class	FTE	Exp	Obs	Rate[b]	Relative Risk[c]	95% CI
Logging operations	9 849	0.93	34	345.20	36.69	25.42, 51.22
Roof work: all types	4 358	0.41	7	160.61	17.07	6.83, 35.17
Garbage collecting	3 932	0.37	5	127.16	13.51	4.36, 31.52
Street and road construction	6 003	0.57	7	116.62	12.39	4.96, 25.52
Trucking: interstate and intrastate	15 898	1.50	15	94.35	10.03	5.61, 16.52
Farms: dairy	7 191	0.68	6	83.44	8.87	3.24, 19.29
Interior finished carpentry	4 829	0.45	4	82.83	8.80	2.37, 22.56
Building construction: wood frame	36 701	3.45	25	68.12	7.24	4.69, 10.69
Electric wiring in building	17 755	1.67	5	28.16	2.99	0.97, 6.98
Fruit packing	22 022	2.07	4	18.16	1.93	0.52, 4.94
State government	23 983	2.26	4	16.68	1.77	0.48, 4.54
Orchards: fruit	46 868	4.41	7	14.94	1.59	0.64, 3.27
Auto and truck sales	30 855	2.90	4	12.96	1.38	0.37, 3.53
Total of all industrial classes	3 198 883	. . .	301	9.41	1.00	. . .
Sales personnel—outside	87 928	8.27	5	5.69	0.60	0.20, 1.41
Class H restaurants	97 684	9.19	4	4.10	0.44	0.12, 1.11
Colleges and universities	118 080	11.11	4	3.39	0.36	0.10, 0.92

Note. FTE = full-time equivalent; Exp = expected; Obs = observed; CI = confidence interval.
[a]Washington Industrial Classification.
[b]Rate expressed as brain injuries per 100 000 FTE/year.
[c]Ordered by highest risk (four or more cases).

Acknowledgments

This work was funded by the Accident and Medical Aid Funds of the State of Washington, Department of Labor and Industries. These research monies are targeted to reducing the incidence of and disability related to occupational injuries and illnesses.

Special thanks are due to Rosa Myers, MSN, for assistance in medical record review and to Melinda Fujiwara for help in preparation of the manuscript.

References

1. Kalsbeek WD, McLaurin RL, Harris BSH, Miller JD. The National Head and Spinal Cord Injury Survey: major findings. *J Neurosurg.* 1980;53(suppl):S19–S31.
2. Kraus JF. Epidemiologic features of injuries to the central nervous system. In: Anderson DW, ed. *Neuroepidemiology: A Tribute to Bruce Shoenberg.* Boston, CRC Press; 1991:333–357.
3. Annegers JF, Grabow JD, Kurland LT, et al. The incidence, cause and secular trends of head trauma in Olmsted County, Minnesota, 1935–1974. *Neurology.* 1980;30:912–919.
4. Edna TH, Cappelen J. Hospital admitted head injury. A prospective study in Trondelag, Norway, 1979–80. *Scand J Soc Med.* 1984;12:7–14.
5. Franklin GM, Haug J, Heyer N, Checkoway H, Peck N. Occupational carpal tunnel syndrome in Washington State, 1984–1988. *Am J Public Health.* 1991;81:741–746.
6. Breslow NE, Day NE. *Statistical Methods in Cancer Research. Vol II. The Design and Analysis of Cohort Studies.* Lyons, France: International Agency for Research on Cancer; 1987:91–103. IARC Scientific Publication 82.
7. Klonoff H, Thompson GB. Epidemiology of head injuries in adults: a pilot study. *Can Med Assoc J.* 1969;100:235–241.
8. Steadman JH, Graham JG. Head injuries: an analysis and follow-up study. *Proc R Soc Med.* 1970;63:23–28.
9. Kerr TA, Kay DWK, Lassman LP. Characteristics of patients, type of accident and mortality in a consecutive series of head injuries admitted to a neurosurgical unit. *Br J Prev Soc Med.* 1971;25:179–185.
10. Jameson KG. Surgical lesions in head injuries: their relative incidence, mortality rates and trends. *Aust NZ J Surg.* 1974;44:241–250.
11. Jennett B, Murray A, MacMillan R, et al. Head injuries in Scottish hospitals. *Lancet.* 1977;2:696–698.
12. MacKenzie EJ, Steinwachs DM, Shankar B. Classifying trauma severity based on hospital discharge diagnoses: validation of an ICD-9CM to AIS 85 conversion table. *Med Care.* 1989;27:412–422.
13. Kraus JF, Fife D. Incidence, external causes, and outcomes of work-related brain injuries in males. *J Occup Med.* 1985;27:757–760.
14. Fife V, Fach G, Hollished W, Boyton W. Incidence and outcome of hospital-treated head injury in Rhode Island. *Am J Public Health.* 1986;76:773–778.
15. Heyer NJ, Franklin G, Rivara FP, et al. Occupational injuries among minors doing farm work in Washington State: 1986 to 1989. *Am J Public Health.* 1992;82:557–560.
16. Tanaka S, Seligman P, Halperin W, Thun M, Timbrook CL, Wasil JJ. Use of workers' compensation claims data for surveillance of cumulative trauma disorders. *J Occup Med.* 1988;30:488–492.

Maternal-Infant HIV Transmission and Circumstances of Delivery

ABSTRACT

Objectives. Circumstances of delivery among children with acquired immunodeficiency syndrome (AIDS) were investigated to assess whether they were consistent with predictions that intrapartum factors affect the risk of maternal-infant human immunodeficiency virus (HIV) transmission.

Methods. Pediatric AIDS patients (maternal-infant transmission; n = 632) reported to the New York City Health Department through 1991 were compared with a series of infants born to predominantly uninfected women. For each case patient, five control subjects were selected and matched from birth certificate files. Hypothesized case–control comparisons for mode of delivery and preselected complications were tested.

Results. Compared with control subjects, case patients were less likely to have been delivered by cesarean section without complications (odds ratio [OR] = 0.77; 95% confidence interval [CI] = 0.59, 1.01) and more likely to have been delivered with complications, whether delivery was by cesarean section (OR = 1.54; 95% CI = 0.98, 2.43) or vaginal (OR = 1.66; 95% CI = 1.15, 2.39).

Conclusions. Assuming that HIV-infected and uninfected women have comparable circumstances of delivery, conditional on sociomedical characteristics, these results suggest that intrapartum events may be associated with maternal-infant HIV transmission. (*Am J Public Health.* 1994;84:1110–1115)

Louise Kuhn, MPH, Zena A. Stein, MBBCH, Pauline A. Thomas, MD, Tejinder Singh, PhD, and Wei-Yann Tsai, PhD

Introduction

Knowledge of the timing of maternal-infant human immunodeficiency virus (HIV) transmission, whether in utero (and if so, at what stage of pregnancy) or at the time of labor and delivery, is critical to both an understanding of the mechanisms of transmission and the development of interventions aimed at preventing infection. Virus has been isolated from fetuses electively aborted early in gestation,[1,2] supporting the inference that at least some transmission takes place in the early intrauterine phase. However, laboratory diagnostic methods (viral culture and polymerase chain reaction) that have excellent sensitivity among older children and adults, if carried out in the neonatal period, identify only half of the children later known to be HIV infected.[3,4]

One explanation for this diagnostic anomaly is that at least half of maternal-infant HIV transmission actually occurs only at parturition; hence, infected cells are scarce at birth and take time to replicate to detectable levels. Further evidence in support of late transmission comes from a study of twins discordant for HIV infection, in which transmission was observed to be less likely for the second-born than for the first-born twin.[5] This finding has been attributed to a briefer exposure to virus in the lower genital tract during delivery of the second twin. Some epidemiological studies have found that babies born by cesarean section have lower rates of infection than babies born vaginally, although none of the individual studies reports statistically significant differences.[6–9] An increased risk of HIV transmission in more complicated deliveries was also observed.[8]

Based on these findings, we hypothesized that if maternal-infant transmission often occurs intrapartum, both mode of delivery and specific complications of delivery that increase or prolong the infant's exposure to maternal blood and secretions at delivery are likely to affect the risk of infection. Unlike other research undertaken to test these relationships, the present study is not based on a cohort of HIV-infected women followed to determine the infection status of their offspring. Such a design is, in principle, to be preferred but is difficult, given the large sample size needed. Instead, we pursue an indirect approach comparing the circumstances of delivery of children with acquired immunodeficiency syndrome (AIDS) reported to the New York City Department of Health with those of a series of uninfected infants born to predominantly uninfected women. If HIV-infected and uninfected women are assumed to be comparable with respect to their circumstances of delivery, conditional on sociomedical characteristics, this comparison should be informative about the association between circumstances of delivery and maternal-infant HIV transmission.

Louise Kuhn and Zena A. Stein are with Columbia University's Division of Epidemiology and Gertrude H. Sergievsky Center, New York, NY. Dr Stein is also with the New York State Psychiatric Institute's HIV Center for Clinical and Behavioral Studies. Pauline A. Thomas and Tejinder Singh are with the New York City Department of Health, AIDS and HIV Surveillance. Wei-Yann Tsai is with Columbia University's Division of Biostatistics.

Requests for reprints should be sent to Louise Kuhn, MPH, Columbia University, Sergievsky Center, P&S Box 16, 630 W 168th St, New York, NY 10032.

This paper was accepted March 2, 1994.

Editor's Note. See related editorial by Goedert and Coté (p 1065) in this issue.

Methods

Of 951 children less than 13 years old diagnosed with AIDS as defined by the Centers for Disease Control and Prevention (maternal-infant transmission) and reported to the New York City Department of Health through 1991, 632 were linked to their New York City birth certificates. Matching criteria included identical or phonetically similar first and last names, identical date of birth, and same sex. Given concordance in other information except the child's last name, a match on the mother's last name was accepted. Failure to link could be owing to clerical or recording errors or name changes, or to births occurring outside New York City. The case patients who could be linked did not differ significantly from unlinked patients in terms of sex, ethnicity, and maternal drug use, but the former were more likely than the latter to have been diagnosed with AIDS before 1 year of age ($\chi^2 = 7.77$, $P = .005$) and less likely to still be alive ($\chi^2 = 3.12$, $P = .077$). This paper reports on the 632 case patients who could be linked, of whom 26% had been diagnosed with AIDS before the age of 1 year, 33% were still alive, and 58% had mothers who were recorded as using injection drugs.

An analysis data set without identifiers was produced combining data from the birth record and the AIDS case report. Data from birth records included mode of delivery (cesarean section or vaginal delivery), complications of labor and delivery, and obstetrical procedures. The delivery complications recorded on the birth certificate were reviewed to identify those likely to increase the risk of maternal-infant HIV transmission owing to the infant's increased or prolonged exposure to maternal blood and fluids. Based on this a priori review, the following complications were defined as risky: placenta previa, abruptio placentae, prolonged labor or failure to progress, premature rupture of membranes, excessive bleeding or hemorrhage, and abnormal presentation. Each case patient was categorized as delivered vaginally, by cesarean section, or by other/unknown mode, with or without the concurrent complications considered risky. A gradient was thus established in which vaginal deliveries without complications were considered the referent category. The study hypothesis predicted that children with AIDS would be the products of fewer cesarean sections without complications and a greater number of vaginal and cesarean deliveries with complications.

The birth certificate records also provided maternal sociodemographic characteristics (ethnicity, age, parity, education, marital status, financial coverage); medical risks during pregnancy (prenatal care, substance use, other medical conditions); and child outcomes (birthweight, gestational age, Apgar scores).

For comparison, five control subjects matched for race, year of birth, and institution of birth were selected from birth certificate files for each case patient. The delivery experience of these control subjects was classified as above, and items of their sociomedical profile were recorded.

The analysis was conducted as follows. First, the sociomedical profile and circumstances of delivery of the case patients and control subjects were compared and tested using chi-square tests. Second, among the control group only, independent associations (Mantel-Haenszel odds ratios [ORs]) between circumstances of delivery and maternal age, parity, education, marital status, prenatal care, financial coverage, drug use during pregnancy, medical risks during pregnancy, gestational age, and birthweight were investigated. Third, circumstances of delivery between both groups were compared within sociomedical strata. Stratum-specific odds ratios were calculated and chi-square tests for trend across the delivery categories within each stratum were performed. Finally, circumstances of delivery between both groups were compared, adjusted for sociomedical characteristics, using multiple logistic regression models. To take into account frequency matching in the data, dummy variables for the matching criteria were added to the logistic models.[10] They were excluded from the final models, however, as they did not change estimates of the regression coefficients.

Results

Maternal and Child Characteristics

A comparison of the maternal and child characteristics of the pediatric AIDS patients and the control group is presented in Table 1. Mothers of children with AIDS were older than comparison mothers and had higher parity. Within maternal age categories, parity remained higher among the case mothers. Case mothers had fewer years of education and were more likely than control mothers to be covered by Medicaid, to have received late or no prenatal care, and to be unmarried. Few of the mothers of either group had received optimal prenatal care, defined as a first visit in the first trimester followed by five or more visits. Case mothers were more likely to have had a sexually transmitted disease during pregnancy and were similarly likely to have had other medical complications during pregnancy, such as diabetes, anemia, preeclampsia, and other conditions. The reported use of substances during pregnancy, especially heroin and cocaine, was higher among mothers of children with AIDS. (A comparison of data on the AIDS case report and birth registry suggests that drug use may be underreported on the birth registry; however, the two data sources are not directly comparable as the birth certificate refers to substance use only during pregnancy. Data from birth certificates would have correctly identified only 29% [107/366] of women recorded on the AIDS case report to inject drugs.)

Children with AIDS had lower birthweights and displayed higher rates of prematurity than control children. Birthweights were lower among case patients independent of gestational age.

Complications and Mode of Delivery

Table 2 presents the frequency among case patients and control subjects of specific risk complications of delivery and mode of delivery. Among case patients, 17.9% were delivered by cesarean section (11.4% were primary and 6.5% were repeat cesarean sections) compared with 19.3% of control subjects (12.2% primary and 7.1% repeat). However, there were fewer cesarean section deliveries without complications among the case patients (13.3%) than among the control group (16.2%), and case patients accounted for an excess of both cesarean and vaginal deliveries with risk complications. Unadjusted for any sociomedical factors, case patients were more likely to have experienced complications during both vaginal (OR = 2.0; 95% confidence interval [CI] = 1.4, 2.8) and cesarean deliveries (OR = 1.5; 95% CI = 1.0, 2.3) and were less likely to have had cesarean deliveries without complications (OR = 0.8; 95% CI = 0.7, 1.1) than control subjects.

From birth certificates, a similar proportion of mothers of AIDS patients (6.6%) and control subjects (6.8%) had had a previous cesarean section. Among the 253 who reported a previous cesarean section, 72.5% of case mothers and 80.3% of control mothers had a cesarean deliv-

TABLE 1—Maternal and Child Characteristics of Pediatric AIDS Patients and Control Subjects, from Birth Certificate Data

	Case Patients, % (n = 632)	Control Subjects, % (n = 3160)	P (χ^2)
Sex: male	52.5	51.5	.621
Ethnicity			
Black	52.7	52.7	
Hispanic	35.9	35.9	
White	11.2	11.2	1.000
Mother's age, y			
≤19	6.5	17.0	
20–29	59.7	57.9	
≥30	33.9	25.0	<.001
Mean	27.3	25.5	
Parity			
1	29.1	44.8	
2	26.4	26.3	
3	20.1	15.3	
4+	24.4	13.6	<.001
Mean	2.6	2.1	
Education			
High school graduate	36.9	40.2	
Any college	9.8	18.0	<.001
Marital status: not married	80.2	60.4	<.001
Finance coverage: Medicaid	68.7	55.1	<.001
Prenatal care			
None	22.2	13.6	
Optimal	16.5	24.1	<.001
Substances used during pregnancy			
Drugs[a]	19.1	2.9	
Tobacco	14.1	7.8	
Alcohol	4.9	1.8	<.001
Medical conditions during pregnancy			
Sexually transmitted disease	4.0	0.8	
Other medical	11.2	9.2	<.001
Birthweight			
≤1499 g	5.1	2.5	
1500–2499 g	23.9	8.8	
≥2500 g	71.0	88.7	<.001
Mean, g	2816	3144	
Gestation			
Preterm	24.7	16.3	
Term	63.5	72.3	
Postterm	11.8	11.4	<.001
Mean, days	268	273	

[a]On the AIDS case reports, 58% of mothers were recorded as using injection drugs.

ery in the index pregnancy (OR = 0.65; 95% CI = 0.28, 1.51).

Predictors of Conditions of Delivery among Control Mothers

Among the control mothers, older age and more education increased the likelihood of having a cesarean delivery. Cesarean sections were more frequent among women who were married, who had received some prenatal care, and who were not covered by Medicaid. Women who reported using drugs or who were recorded as having any other medical complication during pregnancy were also more likely to have had a cesarean delivery. Infants delivered by cesarean section were more likely to register a low birthweight. These associations were independent of complications.

Also among the control group, complications at the time of delivery independent of cesarean section were more frequent among low-birthweight and premature infants. None of the other sociomedical variables was associated with risk complications.

Stratified Analysis

Table 3 compares the conditions of delivery between case patients and control subjects by sociomedical factors. Regardless of sociomedical factor, case patients had a lower risk of being delivered by cesarean section without complications and a greater risk of cesarean or vaginal delivery with complications than control subjects, although confidence intervals are wide. Chi-square tests for trends across the delivery categories were statistically significant in most strata. The estimate of the risk of cesarean deliveries with complications among case patients compared with control subjects was highest among those who had received no prenatal care, but because multiple comparisons were made, interpretation of a single finding like this must be cautious.

Among those women in the study reported to have had a previous cesarean section, a cesarean delivery in the index pregnancy was less likely in the case mothers than in the control mothers. This finding was consistent across all sociomedical strata although statistically significant in none.

From the multiple logistic regression model, adjusted for maternal age, education, marital status, prenatal care, drug use, other medical risks during pregnancy, and Medicaid coverage, case patients were less likely than control subjects to have been delivered by cesarean section without complications (OR = 0.77; 95% CI = 0.59, 1.01) and more likely to have had complications at the time of delivery, whether delivered by cesarean section (OR = 1.54; 95% CI = 0.98, 2.43) or vaginally (OR = 1.66; 95% CI = 1.15, 2.39). These are small associations that reach statistical significance only for vaginal deliveries with complications. When birthweight was included in addition to the confounders listed above, cesarean sections without complications remained less likely but became statistically significant (OR = 0.71; 95% CI = 0.54, 0.94); and complications were still more likely in both cesarean section (OR = 1.33; 95% CI = 0.84, 2.12) and vaginal deliveries (OR = 1.32; 95% CI = 0.91, 1.73) among case patients than among control subjects, although associations for complicated deliveries became weaker.

The inclusion of birthweight in the regression model allows for differences between HIV-infected and uninfected

women to be taken into account in characteristics that may be associated with birthweight, such as drug use; it also makes it possible to control for confounding if birthweight is considered an independent risk factor for transmission. However, it is questionable whether birthweight should be controlled in this analysis. If birthweight is either a cause or consequence of the circumstances of delivery or a marker of intrauterine infection, it is within the causal pathway and the model in which it is not included is more appropriate.

TABLE 2—Mode of Delivery and Risk Complications of Delivery among Pediatric AIDS Patients and Control Subjects

	Case Patients (n = 632)		Control Subjects (n = 3160)		Odds Ratio	95% Confidence Interval
	No.	%	No.	%		
Vaginal deliveries with no risk complications	435	68.8	2262	71.6	1.0	Reference
Cesarean deliveries with no risk complications	84	13.3	512	16.2	0.8	0.7, 1.1
Cesarean deliveries with risk complications	29	4.6	99	3.1	1.5	1.0, 2.3
Abruptio placentae	2	0.3	15	0.5	...	...
Placenta previa	4	0.6	11	0.3	...	...
Bleeding/hemorrhage	1	0.2	0	0.0	...	...
Premature rupture of membranes	6	0.9	20	0.6	...	...
Prolonged labor	10	1.6	30	0.9	...	...
Abnormal presentation	6	0.9	23	0.7	...	...
Vaginal deliveries with risk complications	51	8.1	133	4.2	2.0	1.4, 2.8
Abruptio placentae	5	0.8	6	0.2	...	...
Placenta previa	3	0.5	1	0.0	...	...
Bleeding/hemorrhage	5	0.8	7	0.2	...	...
Premature rupture of membranes	27	4.3	75	2.4	...	...
Prolonged labor	11	1.7	41	1.3	...	...
Abnormal presentation	0	0.0	3	0.1	...	...
Other/unknown	33	5.2	154	4.9	1.1	0.7, 1.6

Discussion

This study suggests that children with AIDS have a pattern of delivery experiences consistent with the hypothesis that events around the time of delivery that increase the infant's exposure to the birth canal and/or maternal blood influence the risk of maternal-infant HIV transmission. Specifically, there is a deficit of uncomplicated cesarean sections and an excess of complications of delivery in the birth histories of children with AIDS. This pattern persists when sociomedical differences between the AIDS patients and the control group with which they are compared are taken into account.

This study is unique in its attempt to consider not only the fact of cesarean delivery but also the obstetric circumstances of the procedure that may affect the risk of transmission. Cesarean sections provide a proxy measure of exposure to maternal blood during delivery; reasons for undertaking a cesarean section, the status of membranes at the time the procedure is undertaken, and other events occurring during surgery are likely to influence the degree of exposure of the infant to maternal blood and fluids. Our inclusion of other complications at delivery improves on previous measurement of this risk and, albeit somewhat crudely, constructs a biological gradient of intrapartum risk.

However, for this study to be informative about mother-infant HIV transmission, we need to make assumptions about the comparability of HIV-infected and uninfected women. Differences between case patients and control subjects may be owing not to the risk of HIV transmission associated with the different intrapartum factors, but instead to differences between HIV-infected and uninfected women. We should note first that, in our analysis, the HIV status of the control mothers is not directly known whereas all of the case mothers are known to be infected. In New York City as a whole, the prevalence of HIV infection among parturient women was 1.16% in 1991, as measured by unlinked serosurveys of newborn infants (heel sticks) by the New York State Department of Health.[11] The prevalence of infection in the control group studied here may be higher, as these subjects are matched to the case patients by race and institution of birth, and HIV-infected births are unevenly distributed among institutions. For practical purposes, however, most of the control subjects can be assumed to be uninfected.

The decision to deliver an infant by cesarean section is based on a combination of social and medical factors. In this paper, we adjust for many of the social factors shown in the data to be associated with having a cesarean section, including maternal age, education, prenatal care, and health insurance. The assumption that HIV-infected and uninfected women are comparable, conditional on these variables, may be considered valid unless other social factors that make HIV-infected women less likely to have a cesarean delivery can be identified. HIV-infected women may be personally and socially disadvantaged or disorganized in ways not captured by the social variables included here. Also, it is not possible to rule out other possibilities—for example, a reluctance of health professionals to perform cesarean sections on women known or suspected to be HIV infected.

Available data strongly suggest that HIV-infected and uninfected women, conditional on certain social variables, are comparable with respect to their rates of cesarean section. In the New York City Perinatal HIV Transmission Collaborative Study, which recruited parturient women with HIV risk behaviors from New York City hospitals, rates of cesarean section are not different between HIV-infected and uninfected women. Among women reporting injection drug use, 22.4% of seropositive women compared with 23.2% of control mothers had a cesarean section ($\chi^2 = 0.01$, $P = .926$); 19.6% of seropositive and 21.7% of seronegative women who did not report injection drug use had a cesarean section ($\chi^2 = 0.37$, $P = .542$).[12] Similarly, other studies in which these data are available do not show statistically different rates of cesarean section among HIV-infected and uninfected women.[13,14]

Even if we assume that our case and control mothers are not comparable in terms of their rates of cesarean section, we would still need to hypothesize addi-

TABLE 3—Differences in Mode of Delivery and Risk Complications[a] between Pediatric AIDS Patients and Control Subjects within Strata of Sociomedical Factors

	Cesarean Section with No Complications		Cesarean Section with Complications		Vaginal Delivery with Complications		P (χ^2 for Trend)
	OR	95% CI	OR	95% CI	OR	95% CI	
Unadjusted	0.85	0.62, 1.17	1.52	0.90, 2.57	1.99	1.33, 2.98	<.001
Mother's age, y							
<30	0.75	0.48, 1.15	1.90	1.01, 3.58	1.74	1.02, 2.95	<.001
≥30	0.90	0.56, 1.46	0.91	0.35, 2.33	2.34	1.26, 4.36	.011
Parity							
Firstborn	0.99	0.57, 1.72	1.57	0.64, 3.85	1.60	0.72, 3.53	.126
Parity ≥2	0.82	0.55, 1.21	1.63	0.85, 3.12	2.30	1.14, 3.67	<.001
Education							
<High school	0.92	0.58, 1.45	1.44	0.62, 3.36	1.39	0.77, 2.52	.109
≥High school	0.82	0.52, 1.31	1.70	0.86, 3.39	2.73	1.55, 4.80	<.001
Marital status							
Married	0.94	0.49, 1.79	1.03	0.29, 3.62	2.15	0.78, 5.87	.138
Unmarried	0.90	0.62, 1.30	1.99	1.12, 3.54	1.82	1.17, 2.82	<.001
Prenatal care							
None	0.84	0.37, 1.86	3.90	1.33, 11.38	1.50	0.67, 3.34	.065
Any	0.86	0.59, 1.27	1.39	0.72, 2.71	2.01	1.18, 3.42	<.001
Medical risks							
Drugs	0.84	0.37, 1.86	1.10	0.26, 4.59	2.07	0.81, 5.28	.075
Other	0.75	0.43, 1.30	1.76	0.72, 4.31	2.34	1.22, 4.48	<.001
None	0.79	0.49, 1.27	1.52	0.73, 3.15	1.63	0.87, 3.03	.007
Payment							
Medicaid	0.87	0.58, 1.30	1.89	1.00, 3.56	2.18	1.39, 3.42	<.001
Other	0.86	0.49, 1.52	1.38	0.54, 3.53	1.47	0.57, 3.77	.140
Birthweight, g							
<2500	0.65	0.36, 1.20	1.18	0.53, 2.63	1.25	0.70, 2.24	.059
≥2500	0.84	0.57, 1.22	1.29	0.64, 2.60	1.65	0.93, 2.94	.009
Gestation							
Term	0.82	0.57, 1.19	1.49	0.79, 2.80	1.68	0.97, 2.89	.002
Preterm	0.94	0.49, 1.83	1.38	0.53, 3.58	2.05	1.11, 3.79	.011

Note. OR = odds ratio; CI = confidence interval.
[a]Vaginal deliveries with no risk complications are defined as the reference group. Subjects in which mode of delivery was missing or defined as "other" are excluded.

tional biases to explain our second finding of an excess of specific complications. None of the sociomedical variables in these data (except birthweight and gestation) was associated with the preselected complications of delivery. Thus, we would have to propose that unidentified disease processes among HIV-infected women increase their risk of complicated deliveries. No statistically significant differences in obstetric complications by serostatus have been reported, although in one study, point estimates of rates of premature rupture of membranes were higher among HIV-positive women (25/80, 31%) than among HIV-negative women (21/109, 19%), but rates of other complications such as placenta previa and abruptio placentae were lower among HIV-positive women (1/85, 1%) than among HIV-negative women (5/118, 4.2%).[15] However, as our estimates suggest modest effects and do not reach statistical significance in every situation, the possibility that our results are owing to small differences between HIV-infected and uninfected women or other unknown biases cannot be definitively ruled out.

From the AIDS case reports, 58% of mothers of children with AIDS are recorded as using injection drugs. It is fairly well established that drug use has an adverse effect on obstetric outcomes,[16] and in the analysis, we control as best we can for maternal drug use during pregnancy. However, this datum may be seriously underreported on the birth certificates, and so adjustment for confounding by drug use may be inadequate. Recording of other medical and risk information on birth certificates may also be incomplete.

Some proportion of infants are likely to be infected before delivery. These cases are uninformative about the risk of intrapartum factors, and were it possible to exclude them, the role of intrapartum factors could be more clearly understood. Reporting of pediatric AIDS in New York City is estimated to be more than 90% complete, but given the relatively recent onset of the epidemic and the long and variable incubation period of AIDS, a cumulative registry tends to favor accrual of cases with short incubation periods. Case patients who could be linked to the birth certificates in this study also appeared to be those diagnosed with AIDS at earlier ages. If the timing of infection affects progression to disease, this series may overrepresent cases attributable to in utero experiences.

If we assume, for the sake of argument, that the results are not due to chance or bias, the measures of effect in this study can be translated into public

health terms. If we assume that the risk of HIV transmission is 25% overall, and use the distribution in the control group to estimate the proportion of specific conditions of delivery while taking the adjusted relative risk estimates from the logistic model, we can calculate the risk of HIV infection under specific conditions of delivery. Under these assumptions, the risk of HIV infection in the group delivered by cesarean section with no complications would be 19%; by cesarean section with complications, 38%; vaginally with complications, 41%; and vaginally with no complications, 25%. Calculating the attributable risk, which is the proportion of cases attributable to the risk factor of interest,[17] 1.6% of cases would be attributable to cesarean sections with complications, 2.6% would be attributable to vaginal deliveries with complications, and 4.6% would be prevented by cesarean sections without complications.

Meta-analysis of several maternal-infant transmission cohorts estimated that 14% of infants delivered by cesarean section are infected compared with 20% of infants delivered vaginally.[18–20] This is a protective effect of a similar order of magnitude to that found in this study for uncomplicated cesarean sections. Further research is needed to clarify the magnitude of the putative protective effect of cesarean sections. Randomized trials would present the strongest design to investigate this as well as to investigate the effect of the procedure on the health of the mother and the child, but such trials may not be ethical or warranted at this time. Further observational evidence is needed that includes data on obstetric circumstances of cesarean sections and potential confounders. As the expected level of protection of an elective cesarean section is modest, the costs and risks associated with the additional surgery may outweigh the potential benefits of the procedure to prevent HIV transmission. Cesarean section is an invasive and expensive intervention[21] associated with small increases in maternal mortality, postoperative and long-term morbidity, and neonatal respiratory problems, all of which are of particular concern in developing countries.[22] Cesarean sections are unlikely to be a successful, widespread means of preventing maternal-infant HIV transmission.

Nevertheless, these findings are consistent with the view that the intrapartum period is a time of risk for HIV transmission. Intrapartum risk has implications for other prevention strategies. The timing of maternal-infant transmission may be important for administration of chemotherapeutic prevention methods, such as nonnucleoside reverse transcriptase inhibitors, other antiretroviral drugs, passive immunization, and, a they become available, other prophylactic medications or vaccines.[23]

As in many other aspects of maternal-child health, prenatal care is critical. Early identification and monitoring of high-risk pregnancies may be beneficial. Unfortunately, many HIV-infected women still do not receive adequate prenatal care. □

Acknowledgments

This study was supported in part by grant 5–P50–MH43520 from the National Institute of Mental Health/National Institute on Drug Abuse.

We would like to thank Dr Mary Ann Chiasson and the New York City Department of Health, Office of Vital Statistics, for collaboration on this project; Dr Shu-Lin Cheng for data abstraction; and Drs Bruce Levin and Ann Gershon for critical review of the study. We would like to acknowledge the work of Sara Back and Diana Sampaio on preliminary aspects of this study.

References

1. Mano H, Chermann JC. Fetal HIV-1 infection of different organs in the second trimester. *AIDS Res Hum Retrovirus.* 1991; 7:83–88.
2. Socéiro R, Rubinstein A, Rashbaum WK, Lyman WD. Maternofetal transmission of AIDS: frequency of human immunodeficiency virus type 1 nucleic acid sequences in human fetal DNA. *J Infect Dis.* 1992;166: 699–703.
3. Burgard M, Mayaux MJ, Blanche S, et al. The use of viral culture and p24 antigen testing to diagnose human immunodeficiency virus infection in neonates. *N Engl J Med.* 1992;327:1192–1197.
4. Krivine A, Firtion G, Cao L, et al. HIV replication during the first weeks of life. *Lancet.* 1992;339:1187–1189.
5. Goedert JJ, Duliege AM, Amos CI, et al. High risk of HIV-1 infection in first-born twins. *Lancet.* 1991;338:1471–1475.
6. European Collaborative Study. Risk factors for mother-to-child transmission of HIV-1. *Lancet.* 1992;339:1007–1012.
7. Gabiano C, Tovo PA, de Martino M, et al. Mother-to-child transmission of HIV-1: risk of infection and correlates of transmission. *Pediatrics.* 1992;90:369–374.
8. Hutto C, Parks WP, Lai S, et al. A hospital-based prospective study of perinatal infection with HIV-1. *J Pediatr.* 1991; 118:347–353.
9. Thomas PA, Weedon J, Krasinski K, et al. Maternal predictors of perinatal HIV transmission. *J Pediatr Infect Dis.* In press.
10. Schlesselman JJ. *Case-Control Studies: Design, Conduct and Analysis.* New York, NY: Oxford University Press; 1982.
11. Newborn Screening Program. Albany, NY: New York State Department of Health.
12. Unpublished data from New York City Perinatal HIV Transmission Collaborative Study. (Principal Investigator: Dr M Heagarty.)
13. Selwyn PA, Schoenbaum EE, Davenny K, et al. Prospective study of human immunodeficiency virus infection and pregnancy outcomes in intravenous drug users. *JAMA.* 1989;261:1289–1294.
14. Butz A, Hutton N, Larson E. Immunoglobulins and growth parameters at birth of infants born to HIV seropositive and seronegative women. *Am J Public Health.* 1991;81:1323–1326.
15. Minkoff HL, Henderson C, Medez H, et al. Pregnancy outcomes among mothers infected with human immunodeficiency virus and uninfected control subjects. *Am J Obstet Gynecol.* 1990;163:1598–1604.
16. Rodriguez EM, Vermund SH. Maternal drug use and pediatric HIV infection: a methodologic review. *Pediatr AIDS HIV Infect: Fetus Adolesc.* 1991;2:107–121.
17. Kleinbaum D, Kupper L, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods.* Belmont, Mass: Lifetime Learning Publications; 1982.
18. Peckham C. Mother-to-child transmission of HIV: risk factors and timing. Ninth International Conference on AIDS. June 6–11, 1993; Berlin, Germany. Abstract PS-04-2.
19. Villari P, Spino C, Chalmers TC, Lau J, Sacks S. Cesarean section to reduce perinatal transmission of HIV. *Online J Curr Clin Trials.* 1993;2(July 8): Document no. 74.
20. Tovo PA. Cesarean section and perinatal transmission: what next? *Lancet.* 1993;342: 630.
21. Rates of cesarean delivery—United States, 1991. *MMWR.* 1993;42:285–289.
22. De Muylder X. Caesarean sections in developing countries: some considerations. *Health Policy Plann.* 1993;8:101–112.
23. Peckham CS, Newell ML, advisory eds. Measures to decrease the risk of mother-to-child transmission of HIV infection. Seminar meeting: January 11–13, 1993; London, England.

Substance Use and HIV-Related Sexual Behaviors among US High School Students: Are They Related?

ABSTRACT

Objectives. This study was undertaken to examine whether use of alcohol, cigarettes, marijuana, cocaine, and other illicit drugs is related to the likelihood of sexual behaviors that increase risk for human immunodeficiency virus (HIV) infection among youth.

Methods. The 1990 national Youth Risk Behavior Survey was used to collect self-reported information about a broad range of health risk behaviors from a representative sample of 11 631 high school students in the United States.

Results. Students who reported no substance use were least likely to report having had sexual intercourse, having had four or more sex partners, and not having used a condom at last sexual intercourse. Adjusted for age, sex, and race/ethnicity, odds ratios for each of these sexual risk behaviors were greatest among students who had used marijuana, cocaine, or other illicit drugs. Students who had used only alcohol or cigarettes had smaller but still significant increases in the likelihood of having had sexual intercourse and of having had four or more sex partners.

Conclusions. HIV prevention programs for youth should recognize that substance use may be an important indicator of risk for HIV infection and acquired immunodeficiency syndrome through its association with unsafe sexual behaviors. (*Am J Public Health.* 1994;84:1116–1120)

Richard Lowry, MD, MS, Deborah Holtzman, PhD, Benedict I. Truman, MD, MPH, Laura Kann, PhD, Janet L. Collins, PhD, and Lloyd J. Kolbe, PhD

Introduction

In response to the acquired immunodeficiency syndrome (AIDS) epidemic, the Centers for Disease Control and Prevention (CDC) was given responsibility for assisting relevant national organizations, state and local educational agencies, and other organizations that serve youth to provide effective health education and other interventions to prevent the spread of human immunodeficiency virus (HIV) among youth.[1] Information about the distribution and interrelatedness of behaviors that increase risk for HIV infection among adolescents is important to the success of these programs.[2,3]

Previous studies have found that adolescents who have had sexual intercourse are more likely than those who have not to use alcohol, cigarettes, marijuana, and other illicit substances.[4–8] (Alcohol and tobacco may be considered illicit substances when used by minors.) Research by Kandel and others[9–17] has suggested a progression of substance use among adolescents, beginning with use of at least one category of substances that are legal for adults (i.e., alcohol and/or cigarettes). Alcohol and cigarettes act as "gateway drugs," with use of one or both typically preceding use of marijuana and other illicit drugs. Few adolescents try illicit drugs other than marijuana without prior use of marijuana. Similar stages appear in comparative studies conducted in other Western societies.[11]

There are, however, limitations to the utility of the gateway theory in predicting determinants of initiation of adolescent substance use (e.g., peer drug use, low self-esteem, psychological distress). Attempts to validate particular sets of important factors necessary to progress from one stage to the next have met with some conflicting and contradictory results.[18] For example, Kaplan[19] found that self-derogation (low self-esteem) was an important predictive influence in the marijuana initiation stage, but this was not found by Jessor and Jessor[20] when studying the same transition point.

While the determinants of progressive involvement in substance use may vary, the actual sequence in which substances are initiated remains consistent. A recent study by Kandel confirms that the developmental pattern of drug involvement identified in the 1970s still characterizes adolescent pathways of drug involvement today.[17] This study extends the current literature and refines the association between substance use and sexual behaviors by examining the relationships between typical stages of adolescent substance use and specific HIV-related sexual behaviors—including sexual intercourse, multiple sex partners, and condom use—among a nationally representative sample of high school students.

Methods

Study Design

As part of the Youth Risk Behavior Surveillance System recently implemented

The authors are with the Division of Adolescent and School Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Ga.

Requests for reprints should be sent to Richard Lowry, MD, MS, Division of Adolescent and School Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Mailstop K-33, 1600 Clifton Rd, NE, Atlanta, GA 30333.

This paper was accepted November 29, 1993.

Editor's Note. See related editorial by Shelov (p 1066) in this issue.

by the CDC to monitor the prevalence of priority health risk behaviors among youth,[21,22] data about substance use, illicit drug injection, sexual intercourse, and other health risk behaviors were collected in April and May 1990 from a representative sample of US high school students. The 1990 national school-based Youth Risk Behavior Survey used a three stage cluster sample design. The respondent universe consisted of all public and private school students in grades 9 through 12 in the 50 states, the District of Columbia, Puerto Rico, and the Virgin Islands. The first-stage sampling frame included 1921 primary sampling units consisting of large counties or groups of smaller, adjacent counties. Fifty primary sampling units were selected from six strata formed on the basis of urbanization and the percentage of Black or Hispanic students in the unit. The units were selected with probability proportional to school enrollment. In the second stage, 174 schools were selected with probability that was proportional to school enrollment but mathematically biased to increase the chance of selecting schools with a substantial number of Black and Hispanic students. In the final stage, one or two classes of a required subject such as English or social studies were randomly selected within each selected school at each grade 9 through 12. All students in selected classes were eligible to participate in the survey.

A weighting factor was applied to each student record to adjust for nonresponse and the oversampling of Black and Hispanic students. The overall weights were scaled so that the weighted count of students per grade was equal to the number of students for that grade. There were no significant differences in school type (public vs private), urbanicity (metropolitan vs nonmetropolitan area), and size of student enrollment between sampled schools that agreed to participate and those that did not. Characteristics of participating and nonparticipating students could not be compared. To improve comparability to other national data sets and future Youth Risk Behavior Survey sampling frames, the single primary sampling unit and six schools selected from Puerto Rico were excluded from this study. The final weighting factors were adjusted to correct for these deletions.

Survey procedures were designed to protect students' privacy and allow for anonymous participation. Parental consent was obtained prior to survey administration. Trained data collectors administered the questionnaire in the classroom.

Data Analyses

The following sexual risk behaviors were examined: (1) ever having had sexual intercourse; (2) having had sexual intercourse with four or more partners; and (3) not having used condoms at last sexual intercourse.

In addition, students were asked if they had ever used alcohol, cigarettes, marijuana, cocaine, and other illicit drugs such as LSD, PCP, methylenedioxymethamphetamine (MDMA), ecstasy, mushrooms, speed, or heroin. They also were asked if they had ever injected any drug not prescribed by a doctor, such as steroids, cocaine, amphetamines, or heroin.

To examine the relationships between substance use and sexual risk behaviors along the stages of adolescent substance use described by Kandel et al.,[9,14,17] respondents were stratified into one of four mutually exclusive groups based on lifetime use of all five substance categories: (1) students who reported never using alcohol, cigarettes, marijuana, cocaine, or other illicit drugs; (2) students who reported ever using alcohol or cigarettes, but never using marijuana, cocaine, or other illicit drugs; (3) students who reported ever using marijuana, but not cocaine or other illicit drugs; and (4) students who reported ever using cocaine or other illicit drugs. Although route of administration (i.e., injection vs noninjection use of illicit drugs) was not explicitly addressed in Kandel's stages of adolescent substance use, it is logical to assume that adolescents who inject illicit drugs have used an illicit substance other than alcohol, cigarettes, and marijuana. Conversely, adolescents who use illicit drugs such as cocaine, amphetamines, and steroids do not necessarily inject them.

Point estimates and 95% confidence intervals (CIs) were calculated using weighted data and statistical programs that accounted for the complex sample design.[23] Odds ratios (ORs), adjusted for age, sex, and race/ethnicity, were calculated using multiple logistic regression. Differences between point estimates were considered significant if the 95% confidence intervals did not overlap.

Results

The school response rate was 73.8%, and the student response rate was 86.8%. A total of 11 631 questionnaires completed at 124 schools were available for analysis.

Nearly 9 of 10 (88.1%; 95% CI = 86.9%, 89.2%) students reported having had a drink of alcohol at least once during their lifetime, and 72.0% (95% CI = 69.9%, 74.2%) reported having taken at least one or two puffs of a cigarette during their lifetime. Fewer students reported ever having used marijuana (31.4%; 95% CI = 28.6%, 34.1%), cocaine (6.6%; 95% CI = 5.7%, 7.5%), and other illicit drugs (13.1%; 95% CI = 11.4%, 14.9%) such as LSD, PCP, MDMA, ecstasy, mushrooms, speed, or heroin.

Sexual risk behaviors and injection drug use were examined along the progression of adolescent substance use identified by Kandel et al[9,14,17]: (1) 8.9% (95% CI = 7.8%, 9.9%) of students reported never having used any substance; (2) 57.9% (95% CI = 55.3%, 60.4%) had used only alcohol or cigarettes; (3) 18.8% (95% CI = 16.7%, 20.9%) had used marijuana, but not cocaine or other illicit drugs; and (4) 14.5% (95% CI = 12.6%, 16.3%) had used, but not necessarily injected, cocaine or other illicit drugs. More than half (54.2%) the students reported ever having had sexual intercourse, and 19.1% reported having had sexual intercourse with four or more partners during their lifetime (Table 1). Among students who reported ever having sexual intercourse, 53.6% did not use a condom the last time they had sexual intercourse. The prevalence of sexual risk behaviors was least among students who reported no substance use; increased among students who used alcohol or cigarettes; and was greatest among students who used marijuana, cocaine, or other illicit drugs. Approximately 1 of 11 (8.9%) students who used cocaine or other illicit drugs reported injection drug use (Table 1).

In addition to potential involvement in injection drug use, students who used cocaine or other illicit drugs were most likely to report the riskiest combination of sexual behaviors. That is, nearly one third (30.2%) of these students have had sexual intercourse with four or more partners and did not use a condom the last time they had sexual intercourse (Table 2).

Previous analyses of these data have presented the prevalence of substance use and sexual risk behaviors separately by demographic characteristics.[24–27] In this analysis, adjusted odds ratios were calculated to measure the association between lifetime substance use and sexual risk behaviors, controlling for age, sex, and

TABLE 1—Percentage of US High School Students Reporting Selected Risk Behaviors for Acquiring HIV Infection, by Lifetime Pattern of Substance Use: Youth Risk Behavior Survey, 1990

Lifetime Pattern of Substance Use	Ever Had Sexual Intercourse		Had Sexual Intercourse with ≥ 4 Partners		Did Not Use Condom at Last Sexual Intercourse[a]		Ever Injected Illicit Drugs[b]	
	%	95% CI	%	95% CI	%	95% CI	%	95% CI
No substance use[c]	17.0	12.5, 21.5	4.2	2.4, 6.0	44.1	35.6, 52.6	...	...
Alcohol/cigarettes[d]	44.2	41.2, 47.1	10.1	8.2, 12.1	48.1	44.7, 51.5	...	...
Marijuana[e]	79.0	74.5, 83.5	30.4	25.7, 35.1	55.6	50.7, 60.5	...	...
Cocaine/other[f]	85.3	82.3, 88.2	45.6	42.6, 48.6	63.7	59.1, 68.3	8.9	6.3, 11.4
Total population	54.2	51.2, 57.3	19.1	17.2, 21.0	53.6	51.3, 56.0	1.5	1.0, 1.9

Note. CI = confidence interval.
[a]Among students who reported ever having had sexual intercourse.
[b]Only among students who (1) answered affirmatively to the question "During your life, have you ever injected (shot up) any drug not prescribed by a doctor such as steroids, cocaine, amphetamines, or heroin?" and (2) answered affirmatively to either of these questions: "During your life, on how many occasions have you used cocaine in any form (including powder, crack, or freebase)?" or "During your life, on how many occasions have you used any other type of illegal drug such as LSD, PCP, MDMA, ecstasy, mushrooms, speed, or heroin?"
[c]Reported never using alcohol, cigarettes, marijuana, cocaine, or other illicit drugs.
[d]Reported ever using alcohol or cigarettes, but never using marijuana, cocaine, or other illicit drugs.
[e]Reported ever using marijuana, but never using cocaine or other illicit drugs.
[f]Reported ever using cocaine or other illicit drugs such as LSD, PCP, MDMA, ecstasy, mushrooms, speed, or heroin.

TABLE 2—Prevalence of Sexual Intercourse among US High School Students, by Condom Use at Last Sexual Intercourse, Number of Sex Partners, and Lifetime Pattern of Substance Use: Youth Risk Behavior Survey, 1990

Lifetime Pattern of Substance Use	Never Had Sexual Intercourse		Ever Had Sexual Intercourse							
			Condom Used Last Time				Condom Not Used Last Time			
			1 to 3 Sex Partners		≥ 4 Sex Partners		1 to 3 Sex Partners		≥ 4 Sex Partners	
	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI
No substance use[a]	83.0	78.5, 87.5	8.1	5.2, 11.0	1.4	0.4, 2.4	4.6	2.6, 6.6	2.8	1.5, 4.1
Alcohol/cigarettes[b]	55.8	52.9, 58.8	17.7	16.4, 19.0	5.2	4.0, 6.4	16.3	14.6, 18.0	4.9	3.9, 6.0
Marijuana[c]	21.0	16.5, 25.5	22.7	19.8, 25.6	12.4	10.7, 14.1	25.9	22.3, 29.4	18.1	14.1, 22.0
Cocaine/other[d]	14.7	11.8, 17.7	15.5	12.8, 18.3	15.3	12.6, 18.1	24.1	19.6, 28.7	30.2	27.1, 33.4
Total population	45.8	42.7, 48.8	17.2	16.3, 18.2	7.9	7.1, 8.8	17.9	16.3, 19.6	11.2	9.7, 12.6

Note. CI = confidence interval.
[a]Reported never using alcohol, cigarettes, marijuana, cocaine, or other illicit drugs.
[b]Reported ever using alcohol or cigarettes, but never using marijuana, cocaine, or other illicit drugs.
[c]Reported ever using marijuana, but never using cocaine or other illicit drugs.
[d]Reported ever using cocaine or other illicit drugs such as LSD, PCP, MDMA, ecstasy, mushrooms, speed, or heroin.

race/ethnicity. Compared with students who reported no substance use, those who used marijuana, cocaine, or other illicit drugs were more likely to report ever having had sexual intercourse, having had four or more sex partners, and not having used a condom at last sexual intercourse. Students who used only alcohol or cigarettes had smaller but still significant increases in the likelihood of ever having had sexual intercourse and having had four or more sex partners, but they were not different from those who reported no substance use in their use of condoms at last sexual intercourse (Table 3).

Discussion

As with other school-based surveys, our findings are likely to underestimate substance use, injection drug use, and sexual risk behaviors among the entire adolescent population since illicit drug use and early sexual intercourse are associated with school dropouts.[8,28–32] Nevertheless, these data suggest that sexual behaviors place a sizable proportion of high school students at risk for HIV infection. More than half of all high school students have already engaged in sexual intercourse, and less than half of those who are sexually active used a condom the last time they had sexual intercourse. Further, the likelihood of these sexual risk behaviors increases as substance use progresses from alcohol and cigarettes to marijuana, cocaine, and other illicit drugs. Approximately one of nine high school students reported both having had four or more sex partners during their lifetime and not having used a condom at last sexual intercourse. Among students who had used cocaine or other illicit drugs, nearly one third reported this combination of sexual risk behaviors. It should be noted that the

percentage of students who have ever had sexual intercourse represents an upper bound on the proportion at risk for HIV infection and AIDS due to sexual behaviors, given that some individuals who have sexual intercourse always use a condom, and some individuals who have sexual intercourse without a condom may be involved in a long-term monogamous relationship with an uninfected partner.

The strong association found in this study between cocaine use and sexual behaviors that increase risk for HIV infection has been noted by others. Several recent studies have examined the sexual behaviors of inner-city, mostly minority youth who use crack,[33–37] an inexpensive and smokable form of cocaine that is highly addictive. Its use by adolescents and young adults has been associated with high-risk sexual behaviors (e.g., exchanging sex for crack or money to buy crack) and HIV infection. Nearly 50% of HIV-infected adolescents seen at an outpatient clinic in New York City reported heavy crack use.[35]

A small proportion of students place themselves at risk for HIV infection through injection drug use. Still, our study suggests that nearly 190 000 (1.5%) of the approximately 12.5 million[38] high school students in the United States have injected cocaine or other illicit drugs. This estimate represents an upper bound on the number of high school students at risk for AIDS due to injection drug use, given that not all injection drug users share needles.

Our study is cross-sectional in design and cannot provide evidence of a causal relationship between substance use and sexual behaviors. Nevertheless, the positive association between stages of adolescent substance use and sexual behaviors that increase risk for HIV infection has important implications. HIV prevention programs for youth should recognize that progressive involvement in substance use may be an important indicator of risk for HIV infection and AIDS, not only through injection drug use but also, and perhaps more importantly, through its association with unsafe sexual behaviors. The association between early stages of substance use and sexual risk behaviors suggests that HIV prevention programs for youth should discourage all substance use. Although cocaine and other illicit drugs were associated with both injection drug use and the greatest likelihood of sexual risk behaviors, use of alcohol and cigarettes were associated with sexual risk behaviors also, albeit to a lesser degree. Further, it is rare for an adolescent to use cocaine or other illicit drugs without first using alcohol, cigarettes, and marijuana.[9,14,17]

TABLE 3—Adjusted Odds Ratios (ORs)[a] and Confidence Intervals (CIs) for Sexual Risk Behaviors, by Lifetime Pattern of Substance Use: Youth Risk Behavior Survey, 1990

Lifetime Pattern of Substance Use	Ever Had Sexual Intercourse		Had Sexual Intercourse with ≥4 Partners		Did Not Use Condom at Last Sexual Intercourse[b]	
	OR	95% CI	OR	95% CI	OR	95% CI
No substance use[c]	1.0	Referent	1.0	Referent	1.0	Referent
Alcohol/cigarettes[d]	4.0	3.0, 5.4	2.7	1.7, 4.2	1.2	0.8, 1.8
Marijuana[e]	17.4	12.9, 23.5	9.2	5.7, 14.9	1.7	1.1, 2.5
Cocaine/other[f]	31.4	22.3, 44.2	26.8	17.3, 41.4	2.3	1.5, 3.7

[a]Adjusted for age, sex, and race/ethnicity.
[b]Among students who reported ever having had sexual intercourse.
[c]Reported never using alcohol, cigarettes, marijuana, cocaine, or other illicit drugs.
[d]Reported ever using alcohol or cigarettes, but never using marijuana, cocaine, or other illicit drugs.
[e]Reported ever using marijuana, but never using cocaine or other illicit drugs.
[f]Reported ever using cocaine or other illicit drugs such as LSD, PCP, MDMA, ecstasy, mushrooms, speed, or heroin.

Another implication of our study involves substance abuse treatment programs that serve youth. Our findings suggest that adolescents who use substances, especially cocaine and other illicit drugs, should receive appropriate counseling and other interventions to reduce sexual behaviors that increase their risk for HIV infection.

National health objectives for the year 2000 seek to increase school-based education to prevent HIV infection (objective 18.10), alcohol and other drug use (objective 4.13), and tobacco use (objective 3.10), preferably as part of quality school health education.[39] There is empirical evidence that well-planned prevention curricula based on social science principles can have a favorable impact on tobacco, alcohol, and other drug use[40] as well as on AIDS-related knowledge, beliefs, self-efficacy, and behaviors among high school students.[41] Substantial and long-lasting reductions in substance use and sexual risk behaviors may require comprehensive, multicomponent interventions that integrate theory-based skill-building education programs for youth with individual counseling, parental involvement and training, and community involvement to ensure access to needed social and health services.[32,42–45] Schools can help by providing developmentally appropriate HIV prevention education within the context of planned and sequential, kindergarten through 12th grade, comprehensive school health education programs that integrate and coordinate efforts to prevent substance use, HIV infection, sexually transmitted diseases, unintended pregnancies, and other important health problems.[46] □

Acknowledgments

Material in this article was previously presented at the 119th Annual Meeting of the American Public Health Association, November 1991, Atlanta, Ga.

References

1. Moore JR, Daily L, Collins J, et al. Progress in efforts to prevent the spread of HIV infection among youth. *Public Health Rep.* 1991;106:678–686.
2. Kann L, Nelson GD, Jones JT, Kolbe LJ. Establishing a system of complementary school-based surveys to annually assess HIV-related knowledge, beliefs, and behaviors among adolescents. *J Sch Health.* 1989;59:55–58.
3. Kolbe LJ. Behavioral epidemiology of AIDS: an organizing model. In: *AIDS Symposia, 1988.* Washington, DC: American Association for the Advancement of Science; 1988:217–223.
4. Zabin LS. The association between smoking and sexual behavior among teens in US contraceptive clinics. *Am J Public Health.* 1984;74:261–263.
5. Zabin LS, Hardy JB, Smith EA, Hirsch MB. Substance use and its relation to sexual activity among inner-city adolescents. *J Adolesc Health Care.* 1986;7:320–331.
6. Mott FL, Haurin RJ. Linkages between sexual activity and alcohol and drug use among American adolescents. *Fam Plann Perspect.* 1988;20:128–136.
7. Keller SE, Bartlett JA, Schleifer SJ, et al. HIV-relevant sexual behavior among a healthy inner-city heterosexual adolescent

population in an endemic area of HIV. *J Adolesc Health.* 1991;12:44–48.
8. Orr DP, Beiter M, Ingersoll G. Premature sexual activity as an indicator of psychosocial risk. *Pediatrics.* 1991;87:141–147.
9. Kandel DB. Stages in adolescent involvement in drug use. *Science.* 1975;190:912–914.
10. Kandel DB, Faust R. Sequence and stages in patterns of adolescent drug use. *Arch Gen Psychiatry.* 1975;32:923–932.
11. Adler I, Kandel DB. Cross-cultural perspectives on the developmental stages in adolescent drug use. *J Stud Alcohol.* 1981;42:701–715.
12. O'Donnell JA, Clayton RR. The stepping-stone hypothesis: a reappraisal. *Chem Depend.* 1982;4:229–241.
13. Donovan JE, Jessor R. Problem drinking and the dimension of involvement with drugs: a Guttman scalogram analysis of adolescent drug use. *Am J Public Health.* 1983;73:543–552.
14. Yamaguchi K, Kandel DB. Patterns of drug use from adolescence to young adulthood: II. sequences of progression. *Am J Public Health.* 1984;74:668–672.
15. Yamaguchi K, Kandel DB. Patterns of drug use from adolescence to young adulthood, III. predictors of progression. *Am J Public Health.* 1984;74:673–681.
16. Welte JW, Barnes GM. Alcohol: the gate-way to other drug use among secondary-school students. *J Youth Adolescence.* 1985;14:487–498.
17. Kandel D, Yamaguchi K. From beer to crack: developmental patterns of drug involvement. *Am J Public Health.* 1993;83:851–855.
18. Newcomb MD, Maddahian E, Bentler PM. Risk factors for drug use among adolescents: concurrent and longitudinal analyses. *Am J Public Health.* 1986;76:525–531.
19. Kaplan HB. Increase in self-rejection as an antecedent of deviant responses. *J Youth Adolescence.* 1975;4:438–458.
20. Jessor R, Jessor SL. Theory testing in longitudinal research on marijuana use. In: Kandel DB, ed. *Longitudinal Research on Drug Use: Empirical Findings and Methodological Issues.* Washington, DC: Hemisphere; 1978:41–71.
21. Kolbe LJ. An epidemiological surveillance system to monitor the prevalence of youth behaviors that most affect health. *Health Educ.* 1990;21:44–48.
22. Kolbe LJ, Kann L, Collins JL. Overview of the Youth Risk Behavior Surveillance System. *Public Health Rep.* 1993;108(suppl 1):2–10.
23. Shah BV. *SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data.* Research Triangle Park, NC: Research Triangle Institute; April 1981.
24. Centers for Disease Control. Tobacco use among high school students—United States, 1990. *MMWR.* 1991;40:617–619.
25. Centers for Disease Control. Alcohol and other drug use among high school students—United States, 1990. *MMWR.* 1991; 40:776–777, 783–784.
26. Centers for Disease Control. Sexual behavior among high school students—United States, 1990. *MMWR.* 1992;40:885–888.
27. Centers for Disease Control. Selected behaviors that increase risk for HIV infection among high school students—United States, 1990. *MMWR.* 1992;41:231, 237–240.
28. Jessor R, Jessor SL. *Problem Behavior and Psychosocial Development: A Longitudinal Study.* New York, NY: Academic Press; 1977.
29. Mensch BS, Kandel DB. Dropping out of high school and drug involvement. *Sociol Educ.* 1981;61:95–113.
30. Jessor R, Costa F, Jessor SL, Donovan JE. The time of first intercourse: a prospective study. *J Pers Soc Psychol.* 1983;44:608–626.
31. Kolbe LJ, Green L, Foreyt J, et al. Appropriate functions of health education in schools: improving health and cognitive performance. In: Krasnegor NA, Arasteh JD, Cataldo MF, eds. *Child Health Behavior: A Behavioral Pediatrics Perspective.* New York, NY: Wiley; 1986:171–216.
32. Dryfoos J. *Adolescents at Risk: Prevalence and Prevention.* New York, NY: Oxford University Press; 1990.
33. Fullilove MT, Fullilove RE. Intersecting epidemics: Black teen crack use and sexually transmitted disease. *J Am Med Wom Assoc.* 1989;44:146–153.
34. Fullilove RE, Fullilove MT, Bowser BP, Gross SA. Risk of sexually transmitted disease among black adolescent crack users in Oakland and San Francisco, Calif. *JAMA.* 1990;263:851–855.
35. Kipke MD, Futterman D, Hein K. HIV infection and AIDS during adolescence. *Med Clin North Am.* 1990;74:1149–1167.
36. Des Jarlais DC, Abdul-Quader A, Minkoff H, Hoesberg B, Landesman S, Tross S. Crack use and multiple AIDS risk behaviors. *J AIDS.* 1991;4:446–447.
37. Edlin B, Irwin KL, Ludwig DD, et al. HIV infection and crack cocaine smoking in street-recruited urban youth, U.S.A. Presentation at the Eight International Conference on AIDS; July 22, 1992; Amsterdam, the Netherlands.
38. School enrollment—social and economic characteristics of students: October 1990. *Current Population Reports.* Series P-20, no. 460. Washington, DC: US Bureau of the Census; 1990.
39. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services, Public Health Service; 1991. DHHS publication PHS 91–50212.
40. Errecart MT, Walberg HJ, Ross JG, Gold RS, Fiedler JL, Kolbe LJ. Effectiveness of teenage health teaching modules. *J Sch Health.* 1991;61:26–30.
41. Walter HJ, Vaughan MS. AIDS risk reduction among a multiethnic sample of urban high school students. *JAMA.* 1993; 270:725–730.
42. Hayes CD, ed. *Risking the Future: Adolescent Sexuality, Pregnancy, and Child-bearing.* Washington, DC: National Academy Press; 1987.
43. *A Decade of Denial: Teens and AIDS in America.* Washington, DC: US House of Representatives, Select Committee on Children, Youth, and Families; 1992.
44. Lavin AT, Shapiro GR, Weill KS. *Creating an Agenda for School-Based Health Promotion: A Review of Selected Reports.* Cambridge, Mass: Harvard School of Public Health; March 1992.
45. Dryfoos JG. Preventing substance use: rethinking strategies. *Am J Public Health.* 1993;83:793–795.
46. *Adolescent Health.* Vol 1: *Summary and Policy Options.* Washington, DC: US Congress, Office of Technology Assessment; 1991.

Longitudinal Tracking of Adolescent Smoking, Physical Activity, and Food Choice Behaviors

ABSTRACT

Objectives. A major assumption underlying youth health promotion has been that physiological risk factors track from childhood into adulthood. However, few studies have systematically examined how behaviors change during adolescence. This paper describes longitudinal tracking of adolescent health behaviors in two Minnesota Heart Health Program communities.

Methods. Beginning in sixth grade (1983), seven annual waves of behavioral measurements were taken from both communities (baseline n = 2376). Self-reported data included smoking behavior, physical activity, and food preferences.

Results. A progressive increase in the change to weekly smoking status was observed across the smoking status categories. As students began to experiment with smoking, they were more likely to either begin to be or remain regular smokers. Tracking of physical activity and food choice variables was also apparent. In nearly all the follow-up periods, the students identified at baseline as measuring high remained high, and those measuring low remained low.

Conclusions. These results indicate that there is evidence of early consolidation and tracking of physical activity, food preference, and smoking behavior. The early consolidation of health behaviors implies that interventions should begin prior to sixth grade, before behavioral patterns are resistant to change. The smoking results suggest that students are experiencing difficulty quitting smoking; thus, youth smoking cessation interventions are warranted. (*Am J Public Health.* 1994;84:1121–1126)

Steven H. Kelder, PhD, MPH, Cheryl L. Perry, PhD, Knut-Inge Klepp, PhD, and Leslie L. Lytle, PhD, RD

Introduction

Efforts directed toward the primary prevention of cardiovascular disease in youth populations have typically focused on the modification of health behaviors such as physical activity, dietary intake, and cigarette smoking. The major assumptions for the rationale behind youth health promotion have been that (1) a certain proportion of children and adolescents are at excess physiological and behavioral risk; (2) physiological risk factors track from childhood into adulthood; (3) the development of physiological risk factors depends largely on the initiation of health-compromising behaviors (such as poor eating and physical activity habits, and cigarette smoking); and (4) primary prevention can be achieved through the modification of behaviors known to be related to physiological risk factors before behavioral patterns are more fully established and resistant to change.[1,2]

Many studies have provided evidence substantiating these assumptions. The relationship between dietary intake, physical activity, cigarette smoking, and cardiovascular disease for adults is well known.[3,4] There is early physiological evidence of cardiovascular disease where fatty streaks and atheromatous changes have been found in the aortas and coronary vessels of children and young adults dying of unrelated disease and war injuries.[5,6] During the past decade there have been six major studies tracking physiological risk factors for cardiovascular disease in childhood, and these studies have each indicated that total, low-density, and high-density lipoprotein cholesterol measurements in childhood and adolescence are predictive of adult values.[7–13] The onset of smoking begins with initial favorable attitudes toward smoking, and as youth move through middle and high school, a substantial proportion progress from initiation to occasional smoking and on to regular nicotine-dependent smoking.[14,15] Few individuals initiate smoking past high school.[16] Finally, several primary prevention studies have investigated various youth health promotion strategies and found encouraging results.[17]

Few studies have systematically examined how cardiovascular disease–related behaviors change during childhood and adolescence within a cohort of youth. If youth health behaviors track during childhood and adolescence, it would add support to the primary assumptions given for early interventions to prevent cardiovascular disease. For this paper, tracking is defined as the stability of health behaviors over time, or the maintenance of relative position in rank of behavior over time.

Methods

Research Design

The Class of 1989 Study annually measured a cohort of students from 6th to

Steven H. Kelder is with the Center for Health Promotion Research and Development, School of Public Health, University of Texas Health Science Center at Houston. Cheryl L. Perry and Leslie L. Lytle are with the Division of Epidemiology at the University of Minnesota's School of Public Health, Minneapolis. Knut-Inge Klepp is with the University of Bergen in Norway.

Requests for reprints should be sent to Steven H. Kelder, PhD, MPH, University of Texas Health Science Center at Houston, School of Public Health, Center for Health Promotion Research and Development, PO Box 20186, Houston, TX 77225.

This paper was accepted February 1, 1994.

Editor's Note. See related editorial by Shelov (p 1066) in this issue.

TABLE 1—Percentage of Students Who Became Weekly Smokers over 1 Year

Grade Change	Never Smoker Becoming Weekly Smoker		Experimenter Becoming Weekly Smoker		Quitter Becoming Weekly Smoker		Weekly Smoker Remaining Weekly Smoker	
	No.	%	No.	%	No.	%	No.	%
Reference community								
6th to 7th	600	3.3	254	13.5	23	34.8	24	49.7
7th to 8th	494	4.1	263	15.3	50	31.9	54	64.6
8th to 9th	344	2.8	259	7.6	61	22.7	83	62.0
9th to 10th	258	3.0	285	16.2	60	45.6	97	54.1
10th to 11th	219	2.7	269	9.3	47	38.2	106	80.0
11th to 12th	131	2.2	148	9.2	22	32.5	62	92.2
Intervention community								
6th to 7th	842	1.1	318	5.1	23	13.0	15	12.0
7th to 8th	709	1.6	422	5.8	45	20.1	35	56.8
8th to 9th	595	2.9	375	11.7	47	22.3	39	60.6
9th to 10th	433	2.1	388	9.4	41	26.3	111	45.5
10th to 11th	388	0.6	375	7.9	71	37.5	88	77.8
11th to 12th	299	2.3	337	8.0	52	21.4	97	79.9

Note. Each cell represents that column's percentage of change to weekly smoker over a 1-year interval. For example, 13.5% of experimental smokers in the reference community became weekly smokers between sixth and seventh grades.

12th grade on a variety of health behaviors. The study was designed to test the efficacy of a school-based cardiovascular disease prevention program embedded in the Minnesota Heart Health Program community campaign to reduce heart disease.[1,18] The research design, interventions, and main outcomes of the study have been described elsewhere.[1,18–22]

Two of the six communities in the Minnesota program were involved in the Class of 1989 Study. All sixth graders enrolled in the public schools in both communities were invited to participate in a baseline survey in April 1983, and that grade cohort was surveyed annually in April until their graduation from high school in 1989. The students in the intervention community participated in 5 years of Minnesota Heart Health Program–sponsored behavioral health programs in school from 6th through 10th grade and also were exposed to the Minnesota program community interventions. The intervention results indicate positive community and school intervention effects on youth smoking, physical activity, and food choice behaviors.

Outcome Evaluation

The measures for cigarette smoking each year included standardized self-reported items.[19,23] Two variables were used: weekly smoking prevalence and a multilevel categorical variable of never smoker, experimental smoker, quitter, and weekly smoker. The measures for physical activity each year were self-reported hours of exercise per week and a physical activity score.[20,24] For these analyses, hours of exercise per week was also divided into four categories: less than 1 hour, 2 to 3 hours, 4 to 5 hours, and greater than 6 hours per week. In the physical activity score, which measured frequency and intensity of regular physical activity, each unit increase means greater frequency and intensity of physical activity. For these analyses, the physical activity score was also divided into tertiles. The measure for dietary intake was self-reported food preferences.[21,24] Students were asked to select from 18 food pairs the "one food they would usually eat when they had the choice"; a unit increase means one more healthier food choice. For these analyses, the food choice score was divided into quintiles.

Analysis Methods

The tracking analyses were accomplished for the physical activity and food choice variables by dividing baseline values into categories (quintiles or tertiles) and computing the subsequent mean values for students originally within those categories. If a mean value within any sixth-grade category maintained a relative position in rank compared with that in the other categories, this was interpreted as evidence of tracking. Using PROC GLM in SAS,[25] a model was fit for each year of follow-up (7th to 12th grades) on the three behavioral physical activity and food choice dependent variables. These analyses were stratified by sex and intervention community, and were adjusted for age, sex, and, in the later years of measurement, school and parental occupation (1986 to 1989 for school; 1987 to 1989 for parental occupation).

This type of tracking analysis is meant to be descriptive; evidence of tracking should be observed by plotting the mean values within sixth-grade categories across time. An additional analysis was performed to determine any differential trends within those categories across time. Using PROC GLM repeated measures in SAS,[25] a model was fit that included a sixth-grade category by time interaction. This analysis approach was used to determine the mean change within categories from 7th to 11th grade on the dependent variable. In this case, a nonsignificant result indicates an equal amount of change within each category and would support the tracking hypothesis. Eleventh grade was selected to assess the change over time in preference to 12th grade because of extensive attrition in 12th grade, particularly in the reference community.

A different analytic approach was taken to assess tracking of cigarette smoking behavior. Because few students are smoking in 6th grade and rapid yearly onset is expected up to 10th grade, the change in smoking status from one year to the next was examined. Using the four-level smoking status variable, the change over a year to weekly smoking from each of the four smoking categories—that is, the proportion of students changing from never smoker, experimental smoker, or quitter to weekly smoker; and those students remaining a weekly smoker (over the year)—was calculated. Again, these analyses were meant to be descriptive, and no statistical testing strategy was used.

Results

Tracking of Smoking Behavior

Table 1 presents data for the adjusted change to weekly smoking from year to year within each of the four smoking categories for both the reference and intervention communities. Significant intervention effects have been reported elsewhere, so these results have been stratified by intervention condition.[19] In

the reference community, the shift from never to weekly smoker from sixth to seventh grade was 3.3%; from seventh to eighth grade, 4.1% of never smokers became weekly smokers, and so on. Likewise, the shift from sixth to seventh grade from experimental smoker to weekly smoker was 13.5%.

A consistent pattern in the change to weekly smoking status can be seen in Table 1 across the smoking status categories. These data indicate a stable and consistent multistage pattern of onset, in which fewer students move from never to weekly smoking than move from never smoking to experimenting, quitting, or remaining weekly smokers. The results are striking for the column of weekly smokers remaining weekly smokers, which indicates that students who smoke are increasingly unlikely to quit as they get older. This observation is corroborated by the quitter column, in which a large proportion of students return to weekly smoking after reporting having quit smoking.

Noticeable differences were observed between the reference and intervention communities, particularly from sixth to seventh and from seventh to eighth grades, where the school smoking intervention occurred. In both of these rows, the change to weekly smoking was lower in the intervention community. In addition, nearly all the cells on the intervention half of the table have a lower weekly smoking prevalence compared with the reference half, some by 10% or more.

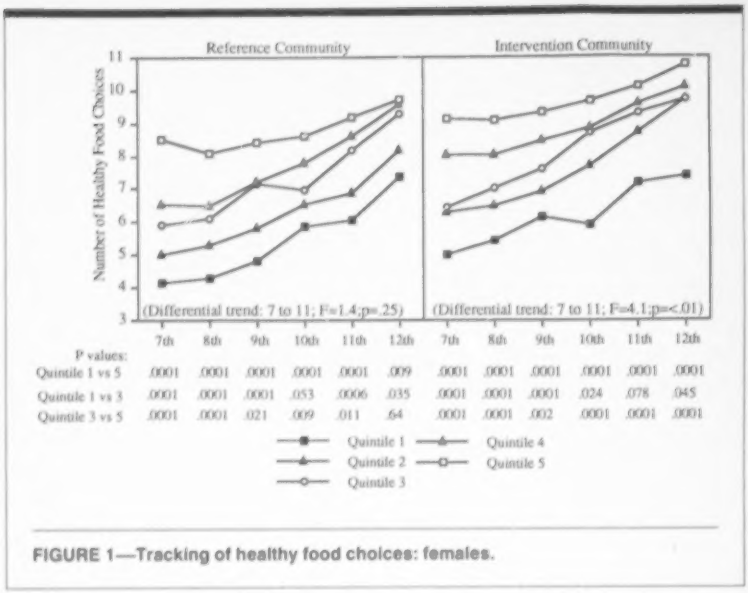

Reference Community
Intervention Community
Number of Healthy Food Choices
11
10
9
8
7
6
5
4
3
(Differential trend: 7 to 11; F=1.4;p=.25)
(Differential trend: 7 to 11; F=4.1;p=<.01)
7th 8th 9th 10th 11th 12th 7th 8th 9th 10th 11th 12th
P values:
Quintile 1 vs 5 .0001 .0001 .0001 .0001 .0001 .009 .0001 .0001 .0001 .0001 .0001 .0001
Quintile 1 vs 3 .0001 .0001 .0001 .053 .0006 .035 .0001 .0001 .0001 .024 .078 .045
Quintile 3 vs 5 .0001 .0001 .021 .009 .011 .64 .0001 .0001 .002 .0001 .0001 .0001
Quintile 1
Quintile 2
Quintile 3
Quintile 4
Quintile 5

FIGURE 1—Tracking of healthy food choices: females.

Tracking of Healthy Food Choices

Figure 1 presents tracking of healthy food choices for females by reference and intervention communities. Significant intervention effects by sex have been reported elsewhere, so results have been stratified by intervention condition and sex.[21] The mean number of healthy food choices was calculated and plotted from 7th to 12th grade within the 6th-grade categories of students. If food choices track, the cohort of 6th-graders' values would be expected to remain separate and distinct over time; if the behavior does not track, the lines would be expected to cross and overlap over time. The bottom portion of Figure 1 includes data on differential trend; this refers to the mean change from 7th to 11th grade within the 6th-grade categories. A nonsignificant test for differential trend is interpreted as no difference in the repeated measures change by 6th-grade category from 7th to 11th grade. Thus, a nonsignificant result supports the tracking hypothesis; a significant differential trend would indicate that at least one category has changed compared with the others over time.

In 6th grade, the healthy food choice variable was divided into quintiles, and the mean number of choices was calculated within those quintiles and plotted from 7th to 12th grade. In Figure 1, a noticeable parallel separation between the groups was observed that visually supports behavioral tracking; that is, sixth-grade rankings of healthy food choices are likely to maintain over time. In sixth grade, at all follow-up periods, those in quintile 1 were significantly different from quintile 5, quintiles 1 and 3 were significantly different at 10 of 12 follow-up periods, and quintiles 3 and 5 were significantly different at 11 of 12 follow-up periods. Additional corroborating evidence is offered by the nonsignificant tests for differential trend in the reference community. A significant differential trend was observed in the intervention community as the increase in the lower two quintiles surpassed that of the upper three, a finding not unexpected since significant intervention effects were observed (reported elsewhere).[21] Similar results were observed for males in that all 12 comparisons between quintiles 1 and 5 and between quintiles 3 and 5 were significant, as were 11 of 12 comparisons between quintiles 1 and 3. Both differential tests for trend were nonsignificant.

Tracking of Activity Score and Hours of Exercise per Week

Figure 2 presents tracking of the activity score (frequency and intensity of physical activity) for sixth-grade females by reference and intervention communities. In these figures, the activity score was divided into tertiles, where tertile 1 represents the least physically active third of the distribution.

A clearly defined pattern of tracking can be observed across the grades. With the exception of one follow-up period, tertile 1 was significantly different from tertile 3. Both tests for differential trend were nonsignificant, corroborating the visual tracking result. The intervention effects of the Class of 1989 Study are also discernible in Figure 2, where the lowest tertile for females appears to have been influenced to a greater extent than tertile 3. This pattern was not unexpected because there were significant intervention effects (reported elsewhere).[21] Similar results were observed for males in that 10 of 12 comparisons between tertiles 1 and 3 were significant, and a significant differential test for trend was observed in the intervention community.

Figure 3 presents tracking of hours of exercise per week for females by reference and intervention communities. Significant intervention effects by sex have been reported elsewhere, so results have

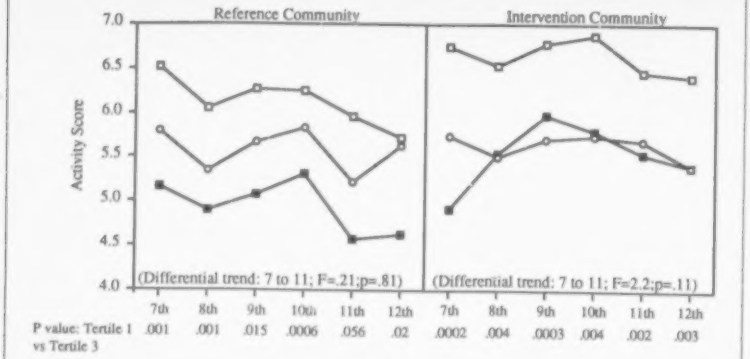

Reference Community
Intervention Community
Activity Score
7.0
6.5
6.0
5.5
5.0
4.5
4.0
(Differential trend: 7 to 11; F=.21;p=.81)
(Differential trend: 7 to 11; F=2.2;p=.11)
7th 8th 9th 10th 11th 12th 7th 8th 9th 10th 11th 12th
P value: Tertile 1 vs Tertile 3 .001 .001 .015 .0006 .056 .02 .0002 .004 .0003 .004 .002 .003

FIGURE 2—Tracking of activity score: females.

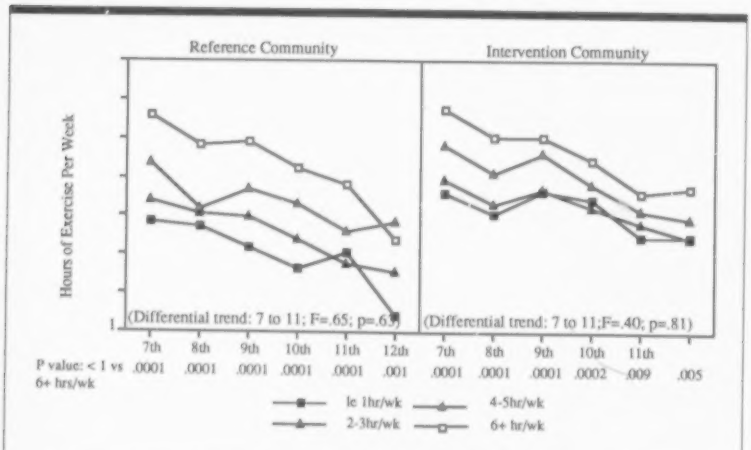

Reference Community
Intervention Community
Hours of Exercise Per Week
(Differential trend: 7 to 11; F=.65; p=.65)
(Differential trend: 7 to 11;F=.40; p=.81)
7th 8th 9th 10th 11th 12th 7th 8th 9th 10th 11th
P value: < 1 vs 6+ hrs/wk .0001 .0001 .0001 .0001 .0001 .001 .0001 .0001 .0001 .0002 .009 .005
lc 1hr/wk
4-5hr/wk
2-3hr/wk
6+ hr/wk

FIGURE 3—Tracking of hours of exercise per week: females.

been stratified by intervention condition and sex.[20] Here again, a clear pattern of tracking can be observed. This observation is most apparent when comparing the highest (6+) with the lowest (<1) groups of exercisers; at each follow-up period, these groups were significantly different. This finding is corroborated by the nonsignificant test for trend in both communities. Similar results were observed for males in that all 12 comparisons between high and low groups were significant, and a significant differential test for trend was observed in the intervention community.

Attrition Analysis

Financial constraints precluded any attempt to contact those students who were not present in class the day of the survey. Response rates for the entire sample for the six follow-up periods were 88%, 81%, 70%, 66%, 59%, and 45% (n's at baseline: intervention = 1342; reference = 1034). These response rates were approximately equal by community with the exception of the last year of data collection, when 55% of the intervention community students were resurveyed compared with 31% of the reference community. Students who were not measured at one survey typically displayed mean values on prior surveys that were less healthy than those of students who were present at more measurement periods.[19-21] Differential attrition bias has been explored and response rates within these categories were nearly equivalent, indicating little if any attrition bias.

Discussion

The concept of behavioral tracking refers to the maintenance of relative ranking of students over time on a given health behavior. These results support the hypothesis that the cardiovascular disease–related behaviors measured in the Class of 1989 Study tracked from 6th to 12th grade. A progressive increase in the change to weekly smoking status was observed across the smoking status categories. A stable and consistent multistage pattern of smoking onset was observed in which fewer students moved to weekly smoking from never smoking than moved from experimenting, quitting, or remaining weekly smokers. Figures 1 through 3, which show the tracking data for the physical activity and food choice variables, all offer visual evidence of tracking. In nearly all the follow-up periods, the students identified in sixth grade as measuring high on any variable remained high, and those measuring low remained low.

These data indicate a clear pattern of tracking and may be interpreted as support for one of the assumptions underlying the rationale for youth health promotion. If students separate at an early age into high- and low-risk groups and maintain their relative ranking over time, presumably into adulthood, then interventions to alter the risk at or before the age of separation are warranted. In addition, these results, coupled with results previously published from these data that indicate a significant amount of covariation among these behaviors,[26] provide evidence that early intervention across multiple risk behaviors is appropriate.

The tracking information not only provides justification for interventions targeting adolescents and younger children, but also may be of assistance to improve the content of future health promotion interventions. A striking pattern of data observed for smoking tracking suggests that once students become weekly smokers, they are unlikely to give up cigarettes. Of the students who were current smokers, an increasing percentage remained smokers over the years of follow-up; they were either unable or unwilling to quit smoking. Of the self-reported quitters, 13% to 46% returned to weekly smoking by the next year's measurement period. Both these observations suggest that nicotine addiction begins early in the students' use of tobacco. From an intervention perspective, these data clearly warrant continued smoking

prevention efforts as well as smoking cessation interventions among high school–aged students. Yet research on the effectiveness of smoking cessation interventions in this age group has not been extensive or encouraging.[27] On the other hand, for the students who self-report having made the effort to quit smoking, 64% to 87% were able to remain a nonsmoker over a year, a rate higher than that reported in most professionally led adult smoking cessation interventions.[28] These higher than expected rates of quitting may be owing to classification error where we cannot be certain of the students' level of involvement with tobacco prior to quitting. Nevertheless, if students who make the effort to quit smoking are relatively successful, encouraging greater numbers of teenage smokers to contemplate or self-initiate quitting may be a useful strategy in combination with smoking prevention programs.

The tracking data for physical activity and healthy food choices offer convincing evidence in support of tracking from 6th to 12th grade. The upward or downward trends across the variables indicate that students are indeed changing their behaviors over time but that the change is relative to the behavior of their peers. This persistence in rankings suggests that consolidation of these health behaviors may begin prior to sixth grade and that interventions in early grade school are warranted. The observation that groups maintain relative ranking indicates that there are subgroups of students who remain at higher risk than their peers. These higher risk groups may require additional intervention to improve their risk profile. Interestingly, the significant tests for differential trend in the intervention community were largely owing to positive changes among those at highest risk. This suggests that the intervention program appears to have had a larger effect among those who needed it most.

Although data from the Class of 1989 Study provide information about patterns of tracking, the limitations must be considered. One possible source of bias for interpretation of the eating and physical activity–related tracking results pertains to measurement. First, the measurement instruments were designed to detect intervention impact and are appropriate for determining differences in group means, but they do not reflect actual nutritional intake and may not accurately reflect true levels of physical activity. There are many difficulties inherent in measuring nutrition and physical activity levels, especially in youth populations.[29] For example, it is clear that healthy food choices are tracking over time, but it is unclear how much these variables relate to actual fat, salt, and carbohydrate consumption. Self-reported measures are subject to demand characteristics such as social desirability or repeated testing that introduces additional within- or between-person variation. The additional variation caused by measurement error, however, would reduce the chances of detecting tracking. Second, self-reported smoking data were not verified with objective biochemical measurements, and these data do not include students who have dropped out of school, a group known to have high rates of smoking.[30] Thus, the smoking rates presented probably underestimate smoking prevalence among adolescents.

The results from the Class of 1989 tracking analysis indicate that there is evidence of early consolidation and tracking of physical activity, food choice, and smoking behaviors. The results have implications for the assumptions underlying school health promotion. First, the early consolidation of health behaviors implies that interventions should begin prior to sixth grade, before the behavioral patterns become more difficult to change. If the tracking of health behaviors is found to be a general phenomenon, it is possible that interventions designed for younger students may move the entire student distribution toward healthier behavioral practices and further improve the impact of interventions at later grades. Future research should replicate these results, beginning with a younger sample of students.

Second, the smoking results indicate that smoking students are addicted to tobacco and are in need of smoking cessation interventions. This appears to be a fertile area of investigation as very few studies have been published on youth smoking cessation programs. In addition, smoking cessation may be facilitated by inserting cessation messages into existing smoking prevention programs, especially in 9th and 10th grades, when the prevalence of smoking warrants greater cessation efforts. □

Acknowledgments

This research was supported by the National Heart, Lung, and Blood Institute, grant RO1-HL 25523.

References

1. Perry CL, Klepp K-I, Sillers C. Community-wide strategies for cardiovascular health: the Minnesota Heart Health Program youth program. *Health Educ Res Theory Pract.* 1989;4:87–101.
2. Perry CL, Stone EJ, Parcel GS, et al. School-based cardiovascular health promotion: the Child and Adolescent Trial for Cardiovascular Health (CATCH). *J School Health.* 1990;60:406–413.
3. Fraser GE. *Preventive Cardiology.* New York, NY: Oxford University Press; 1986.
4. Neaton JD, Wentworth D. Serum cholesterol, blood pressure, cigarette smoking and death from coronary heart disease. *Arch Intern Med.* 1992;152:56–64.
5. Enos WF, Beyer JC, Holmes RH. Pathogenesis of coronary disease in American soldiers killed in Korea. *JAMA.* 1955;158: 912–914.
6. McNamara JJ, Molot MA, Stremple JF, Catting RT. Coronary artery disease in combat casualties in Viet Nam. *JAMA.* 1971;216:1185–1187.
7. Porkka KV, Viikari JSA, Akerblom HK. Tracking of serum HDL-cholesterol and other lipids in children and adolescents: the Cardiovascular Risk in Young Finns Study. *Prev Med.* 1991;20:713–724.
8. Clark WR, Schott HG, Leaverton PE, et al. Tracking of blood lipids and blood pressure in school age children. The Muscatine Study. *Circulation.* 1978;58:626–634.
9. Laskarezewski P, Morrison JA, deGroot I, et al. Lipid and lipoprotein tracking in 108 children over a four year period. *Pediatrics.* 1979;64:584–591.
10. Pagnan A, Ambrosio GB, Vincenzi M, et al. Precursors of atherosclerosis in children: the Cittadella Study. Follow-up and tracking of total serum cholesterol, triglycerides, and blood glucose. *Prev Med.* 1982; 11:381–390.
11. Orchard TJ, Donahue RP, Kuller LH, Hodge PN, Drash AL. Cholesterol screening in childhood: does it predict adult hypercholesterolemia? The Beaver County experience. *J Pediatr.* 1983;103:687–691.
12. Laurer RM, Lee J, Clarke WR. Factors affecting the relationship between childhood and adult cholesterol levels: the Muscatine Study. *Pediatrics.* 1988;82:309–318.
13. American Health Foundation. *Coronary Artery Disease Prevention: Cholesterol, a Pediatric Perspective.* Wynder E, ed. *Prev Med.* 1989;18:323–409. An American Health Foundation Monograph.
14. Cleary PD, Hitchcock JL, Semmer N, Flinchbaugh LJ, Pinney JM. Adolescent smoking: research and health policy. *Milbank Q.* 1988;66:137–171.
15. McCarthy WJ. The cognitive developmental model and other alternatives to the social skills deficit model of smoking onset. In: Bell CS, Battjes R, eds. *Prevention Research: Deterring Drug Abuse among Children and Adolescents.* Rockville, Md: National Institute on Drug Abuse; 1987: 153–169. USDHHS publication ADM 87–1334.
16. Johnston LD, O'Mally PM, Bachman JG. *Drug Use among American High School Seniors, College Students and Young Adults, 1975–1990.* Vol 1: *High School Seniors.* Rockville, Md: National Institute on Drug Abuse; 1991.

17. Stone EJ, Perry CL, Luepker RV. Synthesis of cardiovascular behavioral research for youth health promotion. *Health Educ Q.* 1989;16:155–169.
18. Blackburn H, Luepker RV, Kline FG. The Minnesota Heart Health Program: a research and demonstration project in cardiovascular disease prevention. In: Matarazzo JD, Weiss SM, Herd JA, Miller NE, Weiss SM, eds. *Behavioral Health: A Handbook of Health Enhancement and Disease Prevention.* New York, NY: Wiley Interscience; 1984:1171–1178.
19. Perry CL, Kelder SH, Murray DM, Klepp K-I. Communitywide smoking prevention: long-term outcomes of the Minnesota Heart Health Program. *Am J Public Health.* 1992;82:1210–1216.
20. Kelder SH, Perry CL, Klepp K-I. Community-wide exercise health promotion: outcomes from the Minnesota Heart Health Program and Class of 1989 Study. *J School Health.* 1993;63:218–223.
21. Kelder SH, Perry CL, Trenkner LL, Klepp K-I. Community-wide youth nutrition education: long-term outcomes from the Minnesota Heart Health Program and the Class of 1989 Study. *Health Educ Res Theory Pract.* In press.
22. Mittelmark MB, Luepker RV, Jacobs DR, et al. Community-wide prevention of cardiovascular disease: education strategies of the Minnesota Heart Health Program. *Prev Med.* 1986;15:1–17.
23. Pechacek TF, Murray DM, Luepker RV, Mittelmark MB, Johnson CA, Schultz JM. Measurement of adolescent smoking behavior: rationale and methods. *J Behav Med.* 1984;7:123–140.
24. Perry CL, Klepp K-I, Halper A, et al. Promoting healthy eating and physical activity patterns among adolescents: a pilot study of "Slice of Life." *Health Educ Res: Theory and Prac.* 1987;2:93–103.
25. *SAS/STAT Users Guide.* Vol 2: *GLM-VARCOMP, Version 6.* 4th ed. Cary, NC: SAS Institute Inc; 1989.
26. Lytle LL, Kelder SH, Klepp K-I, Perry CP. Covariance of adolescent health behaviors: the Class of 1989 Study. *Health Educ Res Theory Pract.* In press.
27. Pallonen UE. Smoking cessation in adolescence. Paper presented at the Western Area Conference on Psychosocial Research; January 1987; Los Angeles, Calif.
28. Lichtenstein E, Glasgow RE. Smoking cessation: what have we learned over the last decade? *J Consult Clin Psychol.* 1992;60:518–527.
29. Baranowski T, Simons-Morton BG. Dietary and physical activity assessment in school-aged children: measurement issues. *J School Health.* 1991;61:195–197.
30. Pirie PL, Murray DM, Luepker RV. Smoking prevalence in a cohort of adolescents including absentees, dropouts, and transfers. *Am J Public Health.* 1988;78:176–178.

Planning the 1997 National Health and Nutrition Examination Survey

The National Center for Health Statistics, Centers for Disease Control and Prevention, is soliciting suggestions for topics to be considered in the planning process for the next National Health and Nutrition Examination Survey (NHANES), scheduled for 1997. As in previous NHANES, the primary purpose will be to measure and monitor the health and nutritional status of the US population.

Please send all research suggestions by *October 1, 1994.* Submit specific recommendations, including brief descriptions of the suggestion's public health importance and the available data collection methodologies, to Vicki Burt, Chief, Survey Planning and Development Branch, Division of Health Examination Statistics, National Center for Health Statistics, Presidential Building, Rm 1000, 6525 Belcrest Rd, Hyattsville, MD 20782.

Low Birthweight at Term and the Timing of Fetal Exposure to Maternal Smoking

Ellice Lieberman, MD, DrPH, Isabelle Gremy, MD, MPH, Janet M. Lang, PhD, ScD, and Amy P. Cohen, BA

ABSTRACT

Objectives. This study was undertaken to evaluate the risk of small-for-gestational-age birth for women who stop smoking or begin to smoke during pregnancy.

Methods. Women with term singleton pregnancies from a hospital-based cohort of 11 177 were classified as (1) nonsmokers; (2) smoked throughout pregnancy; (3) smoked during first trimester only; (4) smoked during first and second trimesters only; and (5) smoked during second and third trimesters or during third trimester only. Risk of small-for-gestational-age birth according to smoking category was estimated and adjusted for confounding factors by logistic regression.

Results. Women who stopped smoking by the third trimester were not at increased risk of small-for-gestational-age birth compared with nonsmokers. Women who began smoking during the second or third trimester had an elevated risk of small-for-gestational-age birth (odds ratio [OR] = 1.83; 95% confidence interval [CI] = 1.25, 2.67) similar to that for women who smoked throughout pregnancy (OR = 2.20; 95% CI = 1.90, 2.54). Risk of small-for-gestational-age birth increased with the number of cigarettes smoked during the third trimester.

Conclusions. It is during the third trimester that smoking retards fetal growth, presenting a compelling opportunity for smoking cessation interventions. Programs must emphasize the importance of not resuming smoking late in pregnancy. (*Am J Public Health.* 1994;84:1127–1131)

Introduction

Women who stop smoking during pregnancy give birth to heavier infants than those who continue to smoke throughout pregnancy.[1–6] However, the few studies investigating the effect of the timing of smoking cessation have not yielded consistent results. Rush and Cassano,[2] as well as Hebel et al.,[5] found that stopping smoking during pregnancy at any time up through the eighth month nullified the negative effect of smoking on birthweight but that stopping in the ninth month did not. In contrast, MacArthur and Knox concluded that women needed to stop by 16 weeks to nullify the effect of smoking, and that stopping between 16 and 30 weeks only moderated the effect on birthweight.[3]

Additionally, most studies report their findings as birthweight differences between nonsmokers, persistent smokers, and quitters[2,3,5] without distinguishing whether those differences resulted from short gestation or impaired fetal growth. As low birthweight in infants of smokers results largely from impaired fetal growth, a more direct measure of effect is the proportion of infants born small for their gestational age.[7] Such infants are known to be at increased risk of fetal mortality as well as of neonatal morbidity and mortality[8–10] (although their risk is not as great as that of infants who are very premature).

How the timing of smoking cessation during pregnancy influences the occurrence of small-for-gestational-age births has important public health implications for the planning of smoking cessation interventions for women of childbearing age. Additionally, it is important to examine the effects for women who begin smoking during pregnancy. This study examines the effect of changes in smoking habits during pregnancy on the occurrence of term undersized births.

Methods

The Delivery Interview Program, conducted at the Boston Hospital for Women from August 1977 to March 1980, was designed to study the association between late pregnancy outcome and a wide variety of exposures. Data collected from interviews and medical record reviews included medical and obstetric history, course of the current pregnancy, and infant outcome, as well as demographic information. Women were also asked about their smoking habits, including the number of cigarettes smoked per day during each trimester of pregnancy. A total of 12 718 women were interviewed and had their medical records reviewed. Further details of the study methodology have been presented elsewhere.[11]

For this analysis, the cohort was limited to the 11 177 women who gave birth to term, liveborn infants of known sex and on whom complete smoking information was available. Only term

Ellice Lieberman and Amy P. Cohen are with the Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Harvard Medical School, Chestnut Hill, Mass. Dr Lieberman and Isabelle Gremy are with the Department of Maternal and Child Health, Harvard School of Public Health. Janet M. Lang is with the Department of Epidemiology and Biostatistics, Boston University School of Public Health.

Requests for reprints should be sent to Ellice Lieberman, MD, DrPH, Department of Obstetrics and Gynecology, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115.

This paper was accepted October 11, 1993.

Editor's Note. See related editorial by Shelov (p 1066) in this issue.

TABLE 1—Percentage Distribution of Demographic and Medical Characteristics among Term Small-for-Gestational-Age (SGA) and Appropriate-for-Gestational-Age (AGA) Infants

	Term SGA, % (n = 1188)	Term AGA, % (n = 9989)	P
Maternal race			
White	66	78	< .001
Black	21	12	
Other	13	10	
Insurance			
Welfare recipient	25	14	< .001
Self-pay	9	9	
Insured	66	77	
Maternal age, y			
< 20	13	7	< .001
20–24	23	16	
≥ 25	64	77	
Marital status			
Single	19	11	< .001
Married	73	84	
Other	8	5	
Education			
< High school graduate	19	12	< .001
High school graduate	26	20	
> High school graduate	55	68	
Began prenatal care after first trimester	14	9	< .001
First pregnancy	57	50	< .001
≥ Two alcoholic drinks/day, first trimester	14	14	NS
Any infant malformation	9	9	NS
Maternal height < 5 ft	14	6	< .001
Maternal body mass index[a] < 18	10	6	< .001
Maternal weight gain, lb			
< 15	14	6	< .001
15–25	42	29	
> 25	44	65	
Pre-eclampsia	5	3	< .001
Urinary tract infection during pregnancy	8	6	< .01
DES exposure	2	1	< .01
Smoking by trimester			
None	56	72	< .001
Throughout	37	20	
First only	4	5	
First and second only	1	1	
Second and third only	3	2	

[a]Weight in kilograms divided by (height in meters)^2
Note. NS = not significant; DES = diethylstilbestrol.

pregnancies (37 or more weeks gestation) were investigated so as to assess the specific effects of smoking on the occurrence of growth retardation rather than of low birthweight. There were not sufficient numbers of preterm infants who were growth retarded to examine the effect of smoking on intrauterine growth in that population. Term infants were categorized as being small for gestational age if their birthweight was less than the 10th percentile for their gestational age and sex using the criteria established by Williams et al.[12] For gestational age, the obstetrician's best estimate at the time of birth, which was available for all women, was used instead of last menstrual period, which was missing for 10% of subjects and was more often missing for women of low socioeconomic status. The two methods were in good agreement, differing by less than 2 weeks for 95% of births.

Women were categorized into one of five groups, according to their smoking habits:

1. Nonsmokers: women who never smoked and those who stopped smoking prior to pregnancy were combined, as both our analyses and the literature indicate no residual effect of smoking prior to pregnancy on infant birthweight[2]
2. Smoked throughout pregnancy
3. Smoked during the first trimester of pregnancy only
4. Smoked during the first and second trimesters of pregnancy only
5. Smoked during the second and third trimesters or during the third trimester only. These women began smoking during either the second trimester (n = 158) or the third trimester (n = 70) and continued to smoke for the remainder of pregnancy. These could represent women who stopped smoking prior to pregnancy and relapsed prior to delivery. Groups starting at different times were combined because of small numbers.

The crude risk of small-for-gestational-age birth according to smoking category was determined. Logistic regression analysis was used to determine the risk for each smoking category while controlling for the following potential confounding factors: maternal education, race, age, marital status, payment source, height, body mass index, pregnancy weight gain, parity, late prenatal care, and medical conditions (pre-eclampsia, urinary tract infection, and history of in utero diethylstilbestrol [DES] exposure). The logistic regression model contained four indicator variables for smoking status (a separate variable for each category of smokers), with nonsmokers as the referent group; the relative risk of small-for-gestational-age birth was determined for each smoking category relative to the nonsmoking group. In all logistic regression analyses, odds ratios (ORs) calculated from regression coefficients were used as estimates of relative risk. Confidence intervals (CIs) were calculated from the standard errors of those coefficients.

Risk of small-for-gestational-age birth could be influenced not only by changes in smoking habits but also by the number of cigarettes smoked. To investigate this, logistic regression was used to estimate the association of such a birth with smoking cessation independently for the two subgroups of smokers for whom there were adequate numbers: light smokers (up to 10 cigarettes per day; n = 1613)

and moderate smokers (11 to 20 cigarettes per day; n = 1098). Those women smoking more than 20 cigarettes per day (n = 428) and those who began smoking during pregnancy (n = 228) were not evaluated because of small numbers. Women were categorized according to first trimester smoking habits, as this was the only time when all the women being evaluated smoked.

In addition to evaluating the occurrence of small-for-gestational-age births, a dichotomous outcome, this study also used multiple linear regression (with birthweight as the dependent variable) to investigate the effect of changes in smoking habits on mean birthweight among term infants. In addition to the confounders noted above, gestational age and infant sex were included in these regression models, as infants of 40 weeks' gestational age are somewhat heavier than 37-week infants and male infants are heavier than female infants. Predicted differences in birthweight and the 95% confidence intervals relative to the nonsmoking group were calculated.

Results

There were 11 177 women with term singleton births included in the current analysis. Within the study population, 1188 (10.6%) women gave birth to small-for-gestational-age infants. Women giving birth to such infants differed in various ways from women whose infants were of appropriate size (Table 1). They were more likely to be Black, on welfare, single, less educated, less than 20 years of age, and primiparous; to begin prenatal care later in pregnancy; and to have medical conditions. Additionally, mothers of undersized infants tended to be shorter and thinner (low body mass index), and to gain less weight during pregnancy. These potential confounders were entered into the multivariate model.

Thirty percent (n = 3367) of the women in this cohort smoked at some time during pregnancy (Table 2). Of the smokers, 2449 (73%) smoked throughout pregnancy, 690 (20%) smoked early in pregnancy but quit, and 228 (7%) did not smoke during the first trimester but began to smoke during the second or third trimester and continued through the end of pregnancy. The frequency of small-for-gestational-age births differed according to smoking category (Table 2). Women who smoked throughout pregnancy were at about twice the risk of giving birth to undersized infants than were nonsmokers (17.7% vs 8.5%). For women who smoked only during the first trimester or during the first and second trimesters and then stopped, the risk was similar to that for nonsmokers (8.7% and 7.2%, respectively, vs 8.5%), while women who began smoking during the second or third trimester of their pregnancy had a rate of small-for-gestational-age birth close to that for women who smoked throughout pregnancy (15.4% vs 17.7%).

These associations did not change in logistic regression analyses that controlled for the potential confounding factors described earlier. The risk of giving birth to an undersized infant for women who smoked during either the first trimester only (OR = 1.04; 95% CI = 0.76, 1.43) or the first and second trimesters (OR = 0.83; 95% CI = 0.43, 1.61) was similar to that for nonsmokers. Women who began smoking during the second or third trimester were at increased risk of such an outcome (OR = 1.83; 95% CI = 1.25, 2.67), as were those who smoked throughout pregnancy (OR = 2.20; 95% CI = 1.90, 2.54). As maternal weight gain could mediate the effect of smoking on fetal growth, an additional analysis was performed without this factor; our results were virtually unchanged.

Women who smoked throughout pregnancy also smoked, on average, more cigarettes per day than those who stopped smoking (Table 2). To disentangle the effects of duration from the effects of quantity, logistic regression analyses were performed (controlling for confounders above) to determine the association of changes in smoking habits and risk of small-for-gestational-age birth separately for light and moderate smokers. Light smokers were those who smoked up to 10 cigarettes per day in the first trimester whereas moderate smokers were those

TABLE 2—Frequency at Term of Small-for-Gestational-Age Births, According to Smoking by Trimester

Smoking by Trimester	No. of Women	Frequency of Small-for-Gestational-Age Births, %	Mean No. of Cigarettes/day
Nonsmoker	7810	8.5	0
Smoked throughout	2449	17.7	15.3[a]
Smoked first trimester only	552	8.7	10.8[a]
Smoked first and second trimesters only	138	7.2	13.5[a]
Smoked second and third or third trimester only	228	15.4	7.7[b]
Total	11 177	10.6	

[a]For comparability, estimates for these groups are based on the number of cigarettes/day smoked during the first timester.
[b]Estimates are based on the number of cigarettes/day smoked during the third trimester, since women in this group did not smoke during the first timester.

TABLE 3—The Risk of Small-for-Gestational-Age Birth at Term for Light and Moderate Smokers, According to Maternal Smoking by Trimester[a]

Infant's Exposure to Smoking, by Trimester	Light Smokers[b] (n = 1613) Odd Ratio	Light Smokers[b] (n = 1613) 95% Confidence Interval	Moderate Smokers[c] (n = 1098) Odd Ratio	Moderate Smokers[c] (n = 1098) 95% Confidence Interval
All trimesters[d]	2.0	1.7, 2.4	2.4	2.0, 3.0
First trimester only	1.0	0.7, 1.5	1.1	0.6, 2.1
First and second trimesters only	0.6	0.2, 1.5	1.2	0.4, 3.6

[a]With nonsmokers (n = 7810) as referent group.
[b]≤ 10 cigarettes/day in first trimester.
[c]11–20 cigarettes/day in first trimester.
[d]Excluding women who smoked only in the third trimester.

TABLE 4—Adjusted Odds Ratio (AORs) and 95% Confidence Intervals (CIs) of Small-for-Gestational-Age (SGA) Birth at Term, According to Number of Cigarettes Smoked during the Third Trimester

Cigarettes/ Day	No. of Women	Percentage of SGA Births	AOR (95% CI)[a]
0	7810	8.5	1.0 (referent group)
1–5	685	14.7	1.69 (1.33, 2.14)
6–10	649	18.0	2.16 (1.72, 2.72)
>10	1343	18.7	2.48 (2.08, 2.95)

[a]From a multiple logistic regression model controlled for education, race, payment source (insured, welfare, other), age, marital status, body mass index, height, weight gain, late prenatal care, parity, exposure to diethylstilbestrol (DES), hypertension, urinary tract infection.

who smoked 11 to 20 cigarettes per day in the first trimester. For both groups, the effect of smoking throughout pregnancy and of smoking cessation were similar to that for the overall cohort. Light smokers had an odds ratio of 1.78 (95% CI = 1.44, 2.20) for the occurrence of undersized birth whereas the odds ratio for moderate smokers was 2.55 (95% CI = 2.10, 3.09). For both groups, women who stopped smoking did not have a significantly increased risk of small-for-gestational-age birth when compared with nonsmokers (Table 3). These data indicate that the effect observed in the overall cohort was not solely owing to the fact that women who stopped smoking were initially lighter smokers. Additionally, while women who started smoking during the second or third trimester smoked fewer cigarettes on average than women who smoked throughout pregnancy, their risk of small-for-gestational-age birth was nearly the same as that for continuous smokers.

To assess the association of changes in smoking habits during pregnancy with infant birthweight, a linear regression analysis was performed. The regression model included as independent variables smoking group and the potential confounding factors noted above, as well as infant sex and gestational age. The model predicted that infant birthweight for women who stopped smoking by the second or third trimester was essentially identical to that predicted for nonsmokers. (Linear regression predicted that infant birthweight for those women who smoked only in the first trimester was 0.8 g lighter than that for nonsmokers, whereas those women who smoked in the first and second trimester had a predicted infant birthweight that was 1.6 g lighter than that for nonsmokers.) Women who smoked throughout pregnancy had infants who were 186 g lighter than those of nonsmokers (95% CI = 170, 202 g) whereas the infants of those who started smoking during the second or third trimester were, on average, 138 g lighter than the infants of nonsmokers (95% CI = 82, 194 g).

Given that smoking seems to act during the third trimester to increase the risk of term undersized birth, the next step was to examine the influence of amount smoked during the third trimester. Third trimester smoking was categorized as none, 1 to 5 cigarettes per day, 6 to 10 cigarettes per day, or more than 10 cigarettes per day. The frequency of small-for-gestational-age births according to quantity smoked during the third trimester (Table 4) indicates an increased risk for heavier smokers (χ^2 for trend, $P < .0001$). A logistic regression analysis that included all the potential confounding factors described above indicates odds ratios for the risk of undersized birth that increase with the amount smoked during the third trimester (Table 4).

Discussion

Our analysis of growth retardation among term infants born to smokers indicates that the risk can be reduced to that for nonsmokers if smoking is stopped before the third trimester, while the risk for those who do not smoke early in pregnancy but smoke in the third trimester will be similar to that for continuous smokers. Results from our analysis concur with those of Rush and Cassano.[2] MacArthur and Knox, on the other hand, found a more limited benefit with regard to infant birthweight for women who stopped smoking between 16 and 30 weeks.[3] In that study, however, women who stopped smoking later in pregnancy (later than 16 weeks) were of lower socioeconomic status than women who stopped earlier. Therefore, failure to adjust for socioeconomic status could account for the lower birthweights noted by these authors among infants of women who stopped smoking later in pregnancy.

Our results, taken together with those of the other studies, suggest that the effect of cigarette smoking on growth retardation occurs during the third trimester of pregnancy. This is also consistent with previous work demonstrating that for babies born early in the third trimester (at 30 to 31 weeks), smoking was not associated with the occurrence of growth retardation.[13]

Finally, our investigation indicates that while any smoking during the third trimester is associated with an increase in undersized birth, the risk also increases with the amount smoked. Studies of smoking and birthweight have not consistently noted a dose response. Rush and Cassano found that infants of heavy smokers (more than 15 cigarettes per day throughout pregnancy) had a crude mean birthweight that was 163 g lighter than infants of women who smoked fewer than 5 cigarettes per day[2]; women smoking 5 to 14 cigarettes had infants with birthweights similar to those born to the highest smoking group. Butler and Goldstein, in investigating smoking after the fourth month of pregnancy, found similar results with women smoking fewer than 5 cigarettes per day giving birth to infants with a mean birthweight that was 120 g lighter than those born to smokers of 20 to 30 cigarettes per day, whereas smokers of 5 to 20 cigarettes had infants with intermediate birthweights.[1] Control for confounders did not alter these findings. Hebel et al. examined birthweight according to amount smoked at the eighth month while controlling for confounding factors.[5] In contrast to our findings and to those cited above, they found no dose response for birthweight according to amount smoked at 8 months of pregnancy.

The potential impact on the occurrence of small-for-gestational-age birth if women did stop smoking during the third trimester is substantial. In this population, 24% of women (n = 2677) smoked during the third trimester and their risk of small-for-gestational age birth was 17.5%. If all the smokers in this cohort had stopped smoking (and their rate of undersized births was that of nonsmokers), the rate of small-for-gestational-age infants in the population overall would have been reduced from 10.6% to 8.5%, a 20%

reduction. This is consistent with the findings of Kleinman and Madans,[14] who estimated that low birthweight would decrease by 19% if all women stopped smoking, and with the findings of Kramer,[15] who estimated that, for a population in which 20% of women smoke, the etiological fraction for cigarette smoking as a cause for intrauterine growth retardation was 22%.

While smoking cessation programs have met with limited success, the evidence that smoking acts during the third trimester to retard fetal growth presents a compelling opportunity for intervention at a time when many women may be motivated to alter their behavior. Such interventions need to emphasize the importance of not relapsing and smoking late in pregnancy. It is also important to emphasize to women that these results should not be interpreted to mean that it is safe to smoke during early pregnancy. Smoking during pregnancy has been associated with a number of adverse pregnancy outcomes other than growth retardation, including placenta previa,[16] premature rupture of membranes,[17] and preterm birth,[18] and it is not clear when during pregnancy smoking acts to increase the risk of these events. Additionally, smoking cessation would result in long-term health benefits for the woman and would decrease her child's exposure to environmental tobacco smoke. □

Acknowledgments

The Delivery Interview Program was supported by a grant from the March of Dimes Birth Defects Foundation.

References

1. Butler NR, Goldstein H, Ross EM. Cigarette smoking in pregnancy: its influence on birth weight and perinatal mortality. *BMJ.* 1972;2:127–130.
2. Rush D, Cassano P. Relationship of cigarette smoking and social class to birth weight and perinatal mortality among all births in Britain, 5–11 April 1970. *J Epidemiol Community Health.* 1983;37:249–255.
3. MacArthur C, Knox EG. Smoking in pregnancy: effects of stopping at different stages. *Br J Obstet Gynaecol.* 1988;95:551–555.
4. Naeye RL. Influence of maternal cigarette smoking during pregnancy on fetal and childhood growth. *Obstet Gynecol.* 1981;57:18–21.
5. Hebel JR, Fox NL, Sexton M. Dose-response of birth weight to various measures of maternal smoking during pregnancy. *J Clin Epidemiol.* 1988;41:483–489.
6. Robitaille Y, Kramer MS. Does participation in prenatal courses lead to heavier babies? *Am J Public Health.* 1985;75:1186–1189.
7. *The Health Benefits of Smoking Cessation.* Washington, DC: US Dept of Health and Human Services, Public Health Service; 1990:367–423. DHHS publication CDC 90–8416.
8. Koops BL, Morgan LJ, Battaglia FC. Neonatal mortality risk in relation to birth weight and gestational age: update. *J Pediatr.* 1982;101:969–977.
9. Kleinman JC, Pierre MB Jr, Madans JH, Land GH, Schramm WF. The effects of maternal smoking on fetal and infant mortality. *Am J Epidemiol.* 1988;127:274–282.
10. Warshaw JB. Intrauterine growth retardation. *Pediatr Rev.* 1986;8:107–114.
11. Lieberman E, Ryan KJ, Monson RR, Schoenbaum SC. Risk factors accounting for racial differences in the rate of premature birth. *N Engl J Med.* 1987;317:743–748.
12. Williams RL, Creasy RK, Cunningham GC, Hawes WE, Norris FD, Tashiro M. Fetal growth and perinatal viability in California. *Obstet Gynecol.* 1982;59:624–632.
13. Lang JM, Cohen A, Lieberman E. Risk factors for small-for-gestational-age birth in a preterm population. *Am J Obstet Gynecol.* 1992;166:1374–1378.
14. Kleinman JC, Madans JH. The effects of maternal smoking, physical stature and educational attainment on the incidence of low birth weight. *Am J Epidemiol.* 1985;121:843–855.
15. Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. *Bull WHO.* 1987;65:663–737.
16. Williams MA, Mittendorf R, Lieberman E, Monson RR, Schoenbaum SC, Genest DR. Cigarette smoking during pregnancy in relation to placenta previa. *Am J Obstet Gynecol.* 1991;165:28–32.
17. Williams MA, Mittendorf R, Stubblefield PG, Lieberman E, Schoenbaum SC, Monson RR. Cigarettes, coffee and preterm premature rupture of the membranes. *Am J Epidemiol.* 1992;135:895–903.
18. McDonald AD, Armstrong BG, Sloan M. Cigarette, alcohol, and coffee consumption and prematurity. *Am J Public Health.* 1992;82:87–90.

Indicators for Monitoring the Growth of Peruvian Infants: Weight and Length Gain vs Attained Weight and Length

Ellen G. Piwoz, ScD, Guillermo Lopez de Romaña, MD, Hilary Creed de Kanashiro, MPhil, Robert E. Black, MD, MPH, and Kenneth H. Brown, MD

ABSTRACT

Objectives. To determine appropriate indicators and age intervals for growth monitoring in this population, the relationship between monthly weight and length indicators and nutritional status at 12 months was evaluated among 102 low-income Peruvian infants.

Methods. Cutoffs for defining adequate vs slow growth were obtained from published reference data, and sensitivities and specificities of each indicator were calculated.

Results. Low weight gain in early infancy, especially from 1 to 2 months, is useful for predicting low weight at 1 year (sensitivity = 81%, specificity = 65%, 25th percentile cutoff). Use of actual weight at 2 months, however, produces comparable results (sensitivity = 86%, specificity = 57%, 50th percentile cutoff). Monthly length gains were weak predictors of low length for age. Neither weight nor length gains accurately predicted the opposite form of undernutrition. Nonhuman milk consumption before 4 months and poor appetite from 3 to 12 months were related to low early weight gain and subsequent undernutrition.

Conclusions. Growth monitoring programs in this population should enroll infants at birth, monitor actual weight, promote exclusive breast-feeding and delayed introduction of nonhuman milk, and provide specific feeding advice to mothers of infants with frequentiy depressed appetites. (*Am J Public Health.* 1994; 84:1132–1138)

Introduction

Monthly growth monitoring has been promoted worldwide as an effective means for detecting growth faltering and preventing undernutrition in infants and young children.[1] These programs are intended to monitor weight gain, rather than attained size, and to provide timely feeding advice when infants fail to grow well.[2] Studies of the impact of growth monitoring have shown mixed results,[3–5] and the debate continues as to whether periodic weighing, charting of growth, and delivery of information on proper infant feeding and care is a wise use of scarce health resources in developing countries.[6–9]

The questions most frequently addressed in evaluations of growth monitoring programs are operational ones: (1) Can health workers be trained to weigh and record infant weights accurately? (2) Do mothers regularly attend the weighing sessions, and do they understand the information contained in the growth charts? (3) Do health workers understand the growth charts, and can they identify growth faltering when it occurs? (4) Do health workers provide nutritional counseling to mothers when it is requested and/or required? (5) Is the advice locally appropriate, compelling, and nutritionally sound? (6) Do mothers understand the advice and act on it? and (7) Does regular weighing with nutrition education improve infant growth and prevent undernutrition?

Failure of growth monitoring to reverse growth faltering may be due to one or more negative responses to the first six questions. In addition, in areas where health infrastructure is extremely poor, growth monitoring may not be effective because distribution and upkeep of scales is not efficient, opportunities for (re)training are limited, and supervision is nonexistent. In such locations, it has been suggested that development of basic health infrastructure should precede the initiation of growth monitoring programs.[6]

On the other hand, it has also been argued that growth monitoring programs may not appear to be effective because high-risk children are not specifically targeted.[10] In many programs, the goal is to regularly weigh all children younger than 2 to 5 years old, and specific attention, such as follow-up home visitation, is not directed at children until growth faltering is detected.[11] Nontargeted programs will appear to have little effect on nutritional status when large numbers of infants and children who are not at risk of undernutrition are included in the denominators of calculations to estimate program impact.

The need to identify sensitive and specific indicators of risk for developing undernutrition that can be easily used in community health programs is apparent, given the difficulty of implementing growth monitoring programs in some locations and the potential benefits of targeting high-risk infants and children within established programs. It would be useful to

Ellen G. Piwoz and Robert E. Black are with the Department of International Health, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Md. Guillermo Lopez de Romaña and Hilary Creed de Kanashiro are with the Instituto de Investigación Nutricional, Lima, Peru. Kenneth H. Brown is with the Program in International Nutrition, Department of Nutrition, University of California, Davis.

Requests for reprints should be sent to Ellen G. Piwoz, ScD, 225 Chancery Rd, Baltimore, MD 21218.

This paper was accepted February 1, 1994.

know, for example, whether growth performance or other characteristics such as infant feeding practices, illness, or appetite history during specific early age periods predict undernutrition in later life. If this is the case, limited growth monitoring (during high-risk ages) could be offered in areas where health infrastructure cannot support comprehensive weighing and educational programs. Growth performance during high-risk periods could also be used to identify children who should be monitored more closely in programs already under way.

The purposes of this paper are to (1) describe monthly growth in a population of low-income Peruvian infants who were followed from birth to 12 months of age, (2) evaluate the sensitivity and specificity of different weight and length indicators for predicting undernutrition at 12 months, and (3) identify infant and family characteristics associated with indicators of high predictive power and with infants who were well nourished at 12 months despite slow growth during high-risk periods. The implications of these findings for growth monitoring programs are discussed.

Methods

Subjects and Field Methods

Data from a community-based longitudinal study of the relationships among nutrition, infection, and growth are used for this analysis. The study was carried out from July 1982 to June 1984 in Pueblo Joven Huáscar, a low-income, periurban community in Lima, Peru. The community and field methods have been described previously.[12–15]

One hundred fifty-six full-term, singleton infants with birth weights of 2500 g or more and no congenital abnormalities were enrolled in the study at birth. Low birthweight was uncommon in this population.[13] Of these infants, 3 dropped out in the first 2 weeks and an additional 34 dropped out during the first year. Those lost to follow-up remained in the study an average of 4.9 months (range = 1–10 months). Failure to complete the study was not related to weight or length gain during the monthly intervals available for study.

Body weights and recumbent lengths were measured by trained anthropometrists within 1 week of birth and at monthly intervals thereafter. Weights were determined with a portable, spring-balanced Salter scale that was sensitive to 100 g. Infants were weighed unclothed and weights were recorded as the mean of two separate, consecutive weighings. Lengths were measured to the nearest 0.1 cm with either a locally constructed length board (in homes) or a Harpenden-Holtain infantometer (in clinics). Periodic standardization exercises were held to minimize measurement error.

Length gain in the first month is not included in this analysis because most birth lengths were measured at local hospitals by multiple observers who were not affiliated with the study. Comparisons found that these measurements were less consistent and precise than the length data obtained by study anthropometrists during the first home visit.

Infants were visited at home thrice weekly to obtain information on their health status on that day and the 2 preceding days. Dietary intake information was obtained once or twice a month by 12-hour daytime observations with food weighing. Breast-milk consumption was measured by the test-weighing method.[16] Procedures for diagnosing and coding symptoms and determining dietary intake have been published previously.[13,14]

Information on usual feeding practices was obtained each month through a questionnaire administered to the infant's primary caretaker. Household sociodemographic information was obtained via questionnaire at the end of the study.

Analytic Methods

Growth increments were calculated in a two-step process. First, each infant's weight and length on his or her monthly birthday (30-day) were estimated by linear interpolation from two successive measurements, up to a maximum of 45 days apart. Increments were calculated as the difference between the interpolated values at the start and end of each 30-day period. Weight or length gain per day can be estimated from the data presented by dividing any monthly increment by 30.

Infants who were not measured within 15 days of their birthdays were assigned a missing value for the month and relevant increments. Although 119 infants completed the study, weight and length at 360 days could be calculated for only 102 infants meeting the criteria described above. Thus, the analyses related to nutritional status at 12 months (360 days) included a maximum of 102 infants.

All monthly comparisons use the maximum number of infants available for study during the time period being examined. Sample sizes vary from month to month, either because the timing of measurements did not meet the interpolation requirements or because of attrition.

Undernutrition at 12 months was defined as weight or length lower than the gender-specific National Center for Health Statistics (NCHS) median for this age[17] by 2 SD or more. These cutoffs are 8.1 kg and 70.7 cm for boys and 7.4 kg and 68.6 cm for girls. Differences in monthly growth among infants who were classified as poorly vs adequately nourished at 12 months, according to either indicator, were evaluated with Student's t test for continuous means.[18(pp54–56)] P values of less than .004 were considered statistically significant. This conservative level was selected by dividing .05 by 12, the number of comparisons performed with the same infants in each analysis.[19] P values between .004 and .05 were considered suggestive of relationships.

Cutoffs for defining adequate vs low gains and attained size were the 5th, 25th, and 50th percentiles of published studies (Guo et al.[20] for the first month, Roche et al.[21] for subsequent months). These cutoffs were used because they were derived from a subset of infants participating in the Fels Longitudinal Study, and Fels data were also used to generate the NCHS reference curves used in growth monitoring programs worldwide. Other percentile cutoffs were examined, but because absolute differences between them were small,[22] the results were not substantially affected and are therefore not presented here.

It is important to note that, for simplicity, the findings are presented in terms of the percentiles of the reference distributions used in the analysis, and not absolute weights or lengths. The 5th, 25th, and 50th percentiles of other studies (e.g., Tanner et al.[23]) do not correspond to the same absolute weight or length values. To avoid confusion among citations, results should be interpreted in terms of the gram or centimeter value being evaluated and not the percentile alone. These values may be obtained from the references cited.

The sensitivity, specificity, and positive and negative predictive values[24(p151)] of absolute size and gains above or below different cutoffs were generated from 2 × 2 tables, as defined in Table 1. The relative risk of developing undernutrition by 12 months, with 95% confidence interval (CI), was calculated for each cutoff by means of Epi-Info's STATCALC procedure.[25] Different indicators were compared by examining relative risks, by summing sensitivities and speci-

TABLE 1—Definitions of Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Relative Risk (RR)

	Nutritional Status at 12 Months		
	Poor (−)	Adequate (+)	Total
Anthropometric value			
Below cutoff (−)	a	b	a + b
Above cutoff (+)	c	d	c + d
Total	a + c	b + d	a + b + c + d

Definitions
- Sensitivity = a/(a + c)
- Specificity = d/(b + d)
- PPV = a/(a + b)
- NPV = d/(c + d)
- RR = a/(a + b) ÷ c/(c + d)

TABLE 2—Peruvian Infants' Average Monthly Weight and Length Gains

	Weight Gain				Length Gain			
Age, mo	n	Mean g	SD	≤0, %[a]	n	Mean cm	SD	≤0, %[a]
0–1	108	820	430	4.6	...[b]	...[b]	...[b]	...[b]
1–2	142	1000	350	1.4	141	3.30	0.82	0.0
2–3	136	800	270	<1.0	136	2.64	0.77	0.0
3–4	133	630	2[illegible]	2.3	134	2.07	0.73	<1.0
4–5	128	490	2[illegible]	5.5	129	1.90	0.80	<1.0
5–6	125	370	[illegible]60	9.6	125	1.55	0.76	<1.0
6–7	120	290	270	16.7	120	1.21	0.68	5.0
7–8	115	250	260	19.1	115	1.26	0.61	1.7
8–9	112	220	270	20.5	110	1.26	0.69	2.7
9–10	108	160	240	29.6	106	0.95	0.57	3.8
10–11	106	190	230	27.4	106	1.01	0.51	4.7
11–12	96	180	280	28.1	96	0.92	0.51	4.2

[a]Percentage of infants with either no gain or a measured loss in weight or length during the interval.
[b]Not calculated. See text.

ficities, and by plotting relative operating characteristic curves (not shown). Relative operating characteristic curves plot sensitivity (y) against 1-specificity (x); the point that is highest from the line $x = y$ is assumed to be the best cutoff and indicator.[26]

Characteristics associated with slow growth (i.e., below the selected cutoff) during critical time periods and adequate size at 12 months despite slow growth during critical periods were identified by using Student's t test to compare the means of continuous variables for infants with low vs adequate gains. Differences in the distributions of categorical variables for both groups were determined with chi-square tests of homogeneity.[18] P values of less than .05 for the first set of comparisons and less than .10 for the second set of comparisons were considered statistically significant. The .10 value was used in the latter comparisons because of the small number of infants in this subgroup analysis (n = 45).

Results

Infants' monthly weight and length gains are presented in Table 2. A total of 27 infants (26.5%) were classified as undernourished according to either their weight (n = 21) or length (n = 18) at 12 months. Twelve infants were low in both weight and length (11.8%), 9 were low in weight only (8.8%), and 6 (5.9%) were low in length but not weight.

Compared with infants with adequate weight for age, infants with low weights at 12 months gained less weight during their first (average difference = 324 g, $P < .006$), second (average difference = 231 g, $P < .003$), fourth (average difference = 168 g, $P < .01$), fifth (average difference = 293 g, $P < .0001$), eighth (average difference = 145 g, $P < .04$), and eleventh (average difference = 137 g, $P < .03$) months and less length during their third (average difference = 0.53 cm, $P < .006$) and ninth (average difference = 0.52 cm, $P < .005$) months. Infants of low length for age gained less weight each month than did infants with adequate length at 12 months, but differences were small and relationships between weight gain and subsequent stunting were not statistically significant during any month. Somewhat surprisingly, length gains were also not significantly different for adequate-length vs stunted infants during any of the intervals studied.

The sensitivity, specificity, and predictive values for different definitions of adequate vs low monthly weight gain from birth to 6 months are presented in Table 3. These intervals were examined because, as shown above, low weight gain in early infancy was consistently related to low weight for age at 12 months. Similar statistics for length gain from 2 to 3 months are also presented because slow linear growth during this period was suggestive of a relationship with low weight for age. Similar comparisons using monthly gains and length for age are not presented because the relationships were weak and inconsistent. Results of these comparisons are mentioned when they are informative.

The results shown in Table 3 indicate that the risk of being low in weight for age at 12 months was greater for infants with low monthly gains from 1 to 5 months according to at least one of the cutoff points presented. For weight gain, the greatest relative risk and optimum sensitivity plus specificity were obtained by monitoring infants from 1 to 2 months of age. Infants who gained less than the reference 25th percentile (n = 45), or 1.08 kg for boys and 890 g for girls, were 5.31 times more likely to be undernourished at 12 months than were infants who gained more than these amounts (n = 54) during the same age period (95% CI = 1.93, 14.65). Weights at 1 month were less but not significantly different for infants who gained less (mean = 4.06 kg ± 574 g) or more (mean = 4.20 kg ± 558 g) than this cutoff. Sensitivities and specificities remained the same when infants were stratified by nutritional status at 1 month with the 25th percentile for weight in Roche et al.[21] used as the cutoff.

To elucidate why the 1- to 2-month age interval is critical, growth performance in previous and subsequent months was compared for all infants with weight gain above (n = 63) and below (n = 79)

the 25th percentile cutoff. This analysis revealed that infants with low gains from 1 to 2 months also tended to gain less weight, on average, from birth to 1 month ($P < .04$), from 2 to 3 months ($P < .04$), from 3 to 4 months ($P < .03$), and, to a lesser extent, from 4 to 5 months of age ($P < .09$). Mean differences ranged from 90 to 175 g. Monthly weight gains after 5 months were not significantly different for the two groups (mean differences ranged from 5 to 65 g). Length gains were lower among these infants from 1 to 2 months ($P < .002$) and from 2 to 3 months ($P < .03$); there were no differences in average length increments after 3 months.

Other variables associated with weight gain from 1 to 2 months are given in Table 4. Mothers of infants with low gains tended to be slightly older ($P < .04$) and more educated ($P < .03$) than mothers whose infants gained adequately. Although total energy intakes per kilogram of body weight were similar for the two groups during the dietary observations for the month, low gainers consumed less total energy from breast milk than did infants who gained adequately ($P < .005$). Furthermore, infants who gained slowly received 76% of their total energy from breast milk, compared with 86% for infants with adequate gains ($P = .06$). Low gainers were more likely to consume nonhuman milk than were infants who gained adequately ($P < .04$). Slow growth from 1 to 2 months was not associated with family income, the construction quality of the infant's home, or with the prevalence of diarrhea, fever, anorexia, or respiratory infections during the first or second month of life.

Identifying the traits of infants who gained slowly from 1 to 2 months but managed to achieve adequate weight status by 12 months ("positive deviants") may be useful for defining the nutrition education component of growth monitoring programs. Infants who gained slowly from 1 to 2 months but were not low in weight for age at 12 months had mothers with more than 5 years of formal education ($P < .03$) and lived in homes of better construction quality, specifically homes with latrines and/or cement floors ($P < .05$). In addition, they gained 71% more weight ($P < .001$) from 3 to 5 months of age (1.22 kg ± 395 g) than did infants who became undernourished (720 g ± 516 g), suggesting that this is another critical period for growth monitoring intervention.

Monthly illness prevalences for the positive deviants and infants who gained

TABLE 3—Weight Gain in Early Infancy and Length Gain from 2 to 3 Months as a Predictor of Low Weight at 12 Months

Age, mo	Cutoff Percentile	Sensitivity, %	Specificity, %	PPV, %	NPV, %	RR	95% CI
			Weight gain				
0–1	5th	33	88	42	84	2.54	1.06, 6.11
	25th	67	72	38	89	3.62	1.38, 9.45
	50th	80	57	32	92	3.89	1.20, 12.65
1–2	5th	24	92	45	82	2.50	1.14, 5.48
	25th	81	65	39	93	5.31	1.93, 14.65
	50th	90	24	24	90	2.56	0.65, 10.12
2–3	5th	35	88	44	84	2.76	1.31, 5.82
	25th	55	68	31	85	2.10	0.97, 4.59
	50th	70	53	27	87	2.15	0.90, 5.13
3–4	5th	53	74	33	87	2.48	1.13, 5.25
	25th	58	68	31	83	2.33	1.03, 5.25
	50th	63	58	27	58	1.98	0.85, 4.60
4–5	5th	68	74	39	90	4.14	1.73, 9.88
	25th	74	58	30	90	3.04	1.19, 7.79
	50th	89	44	28	94	5.10	1.25, 20.80
5–6	5th	56	57	23	85	1.51	0.65, 3.49
	25th	61	40	19	82	1.05	0.45, 2.46
	50th	61	40	19	82	1.05	0.45, 2.46
			Length gain				
2–3	5th	40	83	38	84	2.44	1.15, 5.19
	25th	65	67	33	88	2.81	1.23, 6.41
	50th	70	51	27	87	2.06	0.86, 4.93

Note. Definitions of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and relative risk (RR) are given in Table 1. CI = confidence interval.

TABLE 4—Characteristics Associated with Weight Gain Status from 1 to 2 Months

Characteristic	Weight Gain Status[a]	n	Mean	SD	*P*[b]
Mother's age	Low	51	27.8 y	5.84	.04
	Adequate	65	25.5 y	5.44	. . .
Mother's education	Low	51	6.0 y	3.00	.03
	Adequate	65	4.8 y	3.16	. . .
No latrine in home	Low	51	24%	. . .	.30
	Adequate	65	32%	. . .	. . .
Infant weight at 1 mo	Low	63	4060 g	574	.15
	Adequate	79	4200 g	558	. . .
Average daily total energy intake[c]	Low	48	117 kcal/kg	25	.32
	Adequate	63	113 kcal/kg	16	. . .
Average daily breast-milk energy intake[c]	Low	48	383 kcal	170	.005
	Adequate	63	472 kcal	150	. . .
Nonhuman milk consumed regularly (reported)[d]	Low	61	61%	. . .	.18
	Adequate	79	49%	. . .	. . .
Nonhuman milk consumed (observed)[c]	Low	46	48%	. . .	.04
	Adequate	63	29%	. . .	. . .

[a]See text for definition.
[b]Obtained from Student's *t* tests for continuous variables and chi-square tests of homogeneity with 1 *df* for dichotomous variables.
[c]Obtained from the 12-hour dietary observations during the month (1–2 months).
[d]Obtained from monthly questionnaires (1–2 months).

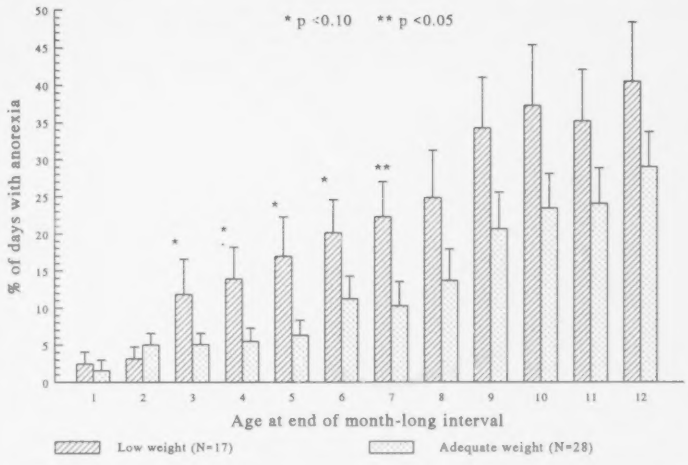


FIGURE 1—Average monthly anorexia prevalence (+1 SE) for infants with low weight gain from 1 to 2 months, by their weight-for-age status at 12 months.

TABLE 5—Attained Weight in Early Infancy as a Predictor of Low Weight at 12 Months

Age, mo	Cutoff Percentile	Sensitivity, %	Specificity, %	PPV, %	NPV, %	RR	95% CI
1	5th	0	97	0	79	. . .[a]	. . .
	25th	29	90	43	83	2.46	1.15, 5.25
	50th	71	71	39	90	4.08	1.73, 9.60
2	5th	29	94	55	83	3.24	1.59, 6.58
	25th	67	80	47	90	4.67	2.10, 10.39
	50th	86	57	35	94	5.54	1.74, 17.63
3	5th	25	96	63	84	3.79	1.87, 7.70
	25th	65	84	50	90	4.67	2.10, 10.39
	50th	90	56	34	96	7.81	1.91, 31.88
4	5th	25	96	63	83	3.75	1.88, 7.61
	25th	70	77	44	91	4.81	2.04, 11.35
	50th	90	60	37	96	9.00	2.21, 36.72
5	5th	50	88	83	97	7.17	3.80, 13.53
	25th	85	79	55	95	11.16	3.52, 35.38
	50th	100	58	38	100	. . .	. . .
6	5th	44	96	73	88	6.18	3.11, 12.27
	25th	83	82	52	96	11.55	3.62, 36.87
	50th	100	51	32	100	. . .	. . .

Note. Definitions of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and relative risk (RR) are given in Table 1.
[a]Cannot be calculated.

slowly at 1 to 2 months and were low in weight for age at 12 months were compared. As shown in Figure 1, the undernourished infants were consistently reported to have had anorexia (reduced appetite) more often from 3 to 12 months than the slow gainers who became well nourished. Although the positive deviant infants suffered from less diarrhea and fever, on average, than the undernourished infants, differences in the prevalences of these illnesses were statistically significant from 8 to 9 months only ($P < .05$). Total energy intake and energy from breast milk, nonhuman milk, and solid foods were not significantly different at any age for the positive deviants vs other infants. The undernourished infants, however, tended to consume more nonhuman milk from 3 to 8 months, and the well-nourished infants tended to consume more total energy and energy from solids from 9 to 12 months of age (data not shown).

Judgments about the use of attained weight vs weight gain for growth monitoring programs can be made by comparing Tables 3 and 5. Although the sensitivities and specificities obtained with attained weight are not, in a single interval, as large as those achieved using weight gain from 1 to 2 months and the 25th percentile until infants are 5 months old (25th percentile; 85% sensitivity, 79% specificity), from 2 months onward the relative risks of becoming undernourished and the sums of sensitivity plus specificity are greater for attained weight than for weight gain with either the 25th or the 50th percentile used as the cutoff.

Discussion

Governments and international agencies have been encouraging the development and implementation of community-based growth monitoring programs as a means of bringing infants and young children into regular contact with health care providers and of identifying growth faltering in its early stages. Evaluations of growth monitoring programs have produced mixed results, leading many to question whether such programs are worthwhile.[3–9]

This analysis was undertaken to evaluate the premise that monthly growth increments are useful predictors of undernutrition at 1 year of age and to identify critical age periods and definitions of inadequate gain for growth monitoring programs. Conditions and behaviors associated with slow growth during high-risk periods were identified.

The study infants' monthly weight gains were not related to their initial birthweights, which were all 2500 g or more. Weight gains during 4 of the first 5 months of life were consistently lower among infants who were identified as low in weight at 12 months than among those who were identified as having adequate weight.

Previous analyses of the patterns of growth in these infants suggested that mean weights were similar to the NCHS reference median until 5 to 6 months of

age[13] and to US breast-fed infants' mean weights until the second half of infancy.[27] Assessment of mean weight alone might lead to the conclusion that growth monitoring is not necessary in this population until infants are 6 months old. This analysis, however, suggests that growth faltering is a cumulative process that begins in early infancy. Weight gain during the second month is a critical period for growth monitoring because infants who gain slowly at this age continue to grow more slowly than other infants until 5 months of age and have a greater risk of being undernourished at 1 year.

Low weight gain in early infancy in this population appears to result from improper infant feeding practices, specifically the early introduction of nonhuman milk to the diet. Compared with infants who gained adequately, infants with low weight gains had older, more educated mothers who were more likely to feed them nonhuman milk in addition to breast milk. However, infants who gained slowly but managed to overcome early growth deficits had mothers with more than a primary-level education and lived in homes of somewhat better socioeconomic status and sanitary environment. For these infants, catch-up growth occurred during the 3- to 5-month period (90–150 days), when those who became well nourished gained 500 g more, on average, than the undernourished.

Infant appetite may have been an important reason why some of the infants who gained slowly in their early months were unable to overcome the growth deficit, whereas others were able to. Infants who gained slowly and remained underweight were reported to have reduced appetites two to three times more often than the "positive deviants" from 3 to 7 months, the age when they should be learning to become accustomed to solid foods. Studies of the epidemiology of anorexia in this population found that reported anorexia was usually preceded or accompanied by other symptoms, such as diarrhea and fever, and 25% to 35% reductions in energy intakes from solid foods.[28] These findings suggest the need to develop appropriate advice on feeding infants with depressed appetites to improve their growth performance.

The comparison of cutoff points and indicators suggests that although monthly weight gain in early infancy is useful for predicting low weight for age, it is not superior to attained weight as an indicator for predicting undernutrition in this population. Neither monthly weight gain nor length gain was accurate for predicting infant stunting.

The failure to find a strong and consistent relationship between monthly length increments and nutritional status at 12 months may have been due to several conditions. Successive monthly length increments were negatively correlated throughout most of infancy, suggesting measurement error, erratic growth, or both. The determinants of stunting in this population and the association between 2-month or 3-month length increments and subsequent stunting are currently being investigated. At present, monthly monitoring of infant length is not recommended.

The results of this analysis suggest that the first 5 months of life are the critical ages for monitoring infant growth and that growth monitoring programs in the population studied should enroll infants at birth or no later than 1 month. However, because weight gain was not superior to attained weight under the strictly supervised conditions of this study, and it is not likely to be better than attained weight in program settings where calculating the difference between two measurements inevitably produces greater error than recording a single weight, growth monitoring programs are advised to use attained weight rather than weight gain in this and similar Peruvian populations. The nutrition education that accompanies growth monitoring programs should focus on exclusive breast-feeding and delaying the introduction of nonhuman milk to improve growth in early infancy. Later on, when infants should be receiving solid food regularly, education should include information and advice on how to feed infants with poor appetites.

Caution is advised in drawing conclusions from this report about how growth monitoring programs should be conducted worldwide. The performance of these anthropometric indicators will vary with the prevalence of low birthweight, feeding practices, and other factors that affect growth patterns in infancy.[29]

In populations with many low-birthweight infants, for example, weight gain may be superior to attained weight, as is suggested in the literature.[2,11] Use of attained weight to target growth monitoring programs in these populations may result in a misuse of resources, because low-birthweight infants who are gaining well will be targeted for intervention while infants whose weights have not yet fallen below the designated weight cutoff but who are growing inadequately will not be reached.

The sensitivity and specificity of these indicators also depend on the outcomes being measured. In this study, the outcomes were low weight or length at 12 months. Results might be different if a different outcome (e.g., mortality) or low weight at a different age (e.g., 2 years) were being predicted. This paper illustrates the types of relatively simple analyses that can be performed to address whether growth monitoring programs, as they are currently promoted, are a worthwhile use of scarce health resources and how such programs can be targeted and thus made more efficient. Additional comparisons, using reliable growth data from other populations, must be carried out before guidelines for growth monitoring programs worldwide can be offered. □

Acknowledgments

These studies were supported primarily by the Diarrheal Disease Control Program of the World Health Organization and by the International Development Research Center. Additional funds were provided by the Office of Nutrition, US Agency for International Development (Dietary Management of Diarrhea Program, Cooperative agreement No. DAN-1010-A-00-5119-00).

The authors express their appreciation to the staff of the Instituto de Investigación Nutricional who collected and organized all of the data, and to Drs Stan Becker, Laura Caulfield, Marie Diener-West, and Jan Peerson for their thoughtful comments on this manuscript.

References

1. Morley D. *See How They Grow: Monitoring Child Growth for Appropriate Health Care in Developing Countries.* Oxford, England: Oxford University Press; 1979.
2. Griffiths M. Growth monitoring: making it a tool for education. *Indian J Pediatr.* 1988;55(suppl 1):59–66.
3. Gerein NM, Ross DA. Is growth monitoring worthwhile? An evaluation of its use in three child health programmes in Zaire. *Soc Sci Med.* 1991;32:667–675.
4. Gopaldas T, Christian PS, Abbi R, Gujral S. Does growth monitoring work as it ought to in countries of low literacy? *J Trop Pediatr.* 1990;36:322–327.
5. Gerein N. Is growth monitoring worthwhile? *Health Policy Plann.* 1988;3:181–194.
6. Nabarro D, Chinnock P. Growth monitoring—inappropriate promotion of an appropriate technology. *Soc Sci Med.* 1988;26:941–948.
7. Gopalan C. Growth monitoring: intermediate technology or expensive luxury? *Lancet.* 1985;2:1337–1338.
8. Briend A, Bari A. Critical assessment of the use of growth monitoring for identifying high risk children in primary health care programmes. *BMJ.* 1989;298:1607–1611.

9. Henry F, Briend A, Cooper E. Targeting nutritional interventions: is there a role for growth monitoring? *Health Policy Plann.* 1989;4:295–300.
10. Gadomski A, Black RE, Mosley WH. Constraints to the potential impact of child survival in developing countries. *Health Policy Plann.* 1990;5:235–245.
11. Hendrata L, Rohde JE. Ten pitfalls of growth monitoring and promotion. *Indian J Pediatr.* 1988;55(suppl):9–15.
12. Lopez de Romaña G, Brown KH, Black RE. Health and growth of infants and young children in Huáscar, Perú. *Ecol Food Nutr.* 1987;19:213–229.
13. Lopez de Romaña G, Brown KH, Black RE, Creed de Kanashiro H. Longitudinal studies of infectious diseases and physical growth of infants in Huáscar, an underprivileged peri-urban community of Lima, Perú. *Am J Epidemiol.* 1989;129:769–784.
14. Creed de Kanashiro H, Brown KH, Lopez de Romaña G, Lopez T, Black RE. Consumption of food and nutrients in Huáscar (Lima), Perú. *Am J Clin Nutr.* 1990;52:995–1004.
15. Brown KH, Stallings RY, Creed de Kanashiro H, Lopez de Romaña G, Black RE. Effects of common illnesses on infants' energy intakes from breastmilk and other foods during longitudinal community-based studies in Huáscar (Lima), Perú. *Am J Clin Nutr.* 1990;52:1005–1013.
16. Brown KH, Black RE, Robertson AD, Akhtar NA, Ahmed G, Becker S. Clinical and field studies of human lactation: methodologic considerations. *Am J Clin Nutr.* 1982;35:745–756.
17. *NCHS Growth Curves for Children, Birth to 18 Years.* Hyattsville, Md: National Center for Health Statistics; 1978.
18. Snedecor GW, Cochran WG. *Statistical Methods.* 7th ed. Ames, Iowa: Iowa State University Press; 1980.
19. Miller RG. *Simultaneous Statistical Inference.* New York, NY: Springer-Verlag; 1981.
20. Guo S, Roche AF, Fomon SJ, et al. Reference data on gains in weight and length during the first two years of life. *J Pediatr.* 1991;119:355–362.
21. Roche AF, Guo S, Moore WM. Weight and recumbent length from 1 to 12 mo of age: reference data from 1-mo increments. *Am J Clin Nutr.* 1989;49:599–607.
22. Piwoz EG, Peerson JM, Brown KH. Potential for misclassification of infants' growth increments by using existing reference data. *Am J Clin Nutr.* 1992;56:58–64.
23. Tanner JM, Whitehouse RH, Takaishi M. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children. *Arch Dis Child.* 1966;41:613–635.
24. Lilienfeld AM, Lilienfeld DE. *Foundations of Epidemiology.* 2nd ed. Oxford, England: Oxford University Press; 1980.
25. *Epi-Info Version 5.0.* Atlanta, Ga: Centers for Disease Control; 1990.
26. Erdreich LS, Lee ET. Use of relative operating characteristic analysis in epidemiology. *Am J Epidemiol.* 1981;114:649–662.
27. Dewey KG, Peerson JM, Heinig MJ, et al. Growth patterns of breastfed infants in affluent (US) and poor (Perú) communities: implications for timing of complementary feeding. *Am J Clin Nutr.* 1992;56:1012–1018.
28. Brown KH, Peerson JM, Lopez de Romaña G, Creed de Kanashiro H, Black RE. Validity and epidemiology of reported poor appetite among Peruvian infants from a low-income, peri-urban community. *FASEB J.* 1992;6(5):A1649. Abstract.
29. Habicht JP, Meyers LD, Brownie C. Indicators for identifying and counting the improperly nourished. *Am J Clin Nutr.* 1982;35:1241–1254.

Call for Papers on Breast-Feeding

The *Journal of Human Lactation*, the official journal of the International Lactation Consultant Association, is issuing a call for papers on the following themes: "Lactation by Women with Physical Challenges," for consideration in the September 1995 issue (submission deadline is *October 1, 1994*); and "Breast-feeding and (Maternal and/or Infant) Sleep Considerations," for consideration in a special issue in 1996 (submission deadline is *January 15, 1995*).

Research papers, commentaries, case series, and single case reports on the above topics will be considered for these special issues. For each submission, send an original manuscript and three copies. To obtain author guidelines, fax a request to (708) 957-8824 or send a self-addressed (stamped if in the United States) envelope to the *JHL* Editorial Office–Dept. JPH, 2240 Willow Rd, Homewood, IL 60430-3221.

Public Health Briefs

Ethnic Differences in Midwife-Attended US Births

Jennifer D. Parker, PhD

ABSTRACT

This study examined US ethnic differences in midwifery care from 1982 through 1989. After adjustment for maternal characteristics, Native American mothers were most likely, and White and Asian mothers were least likely, to obtain midwifery care. For these three groups, midwifery use increased rapidly in the period from 1982 through 1989. Compared with White mothers, Black and Hispanic mothers were more likely to be attended by a midwife; however, their use of midwives increased more slowly. Though the local availability of obstetricians and midwives may determine a woman's choice of care provider, these data suggest that cultural factors also play a role. (*Am J Public Health*. 1994;84:1139–1141)

Introduction

In 1989 nearly 150 000 US births were attended by midwives,[1] up from 60 000 a decade earlier.[2] Overall, women served by midwives differ from the general population of obstetric patients[2,3]; compared with all US births in 1988, for example, midwife-attended hospital births (87% of all midwife-attended births) were more common among women who were non-White, younger, less educated, and unmarried.[3] The use of midwives for perinatal care has been found to have no adverse effects on birth outcomes,[5–11] in spite of the dominance of women at higher risk for poor pregnancy outcomes among midwife patients. Because incorporating midwives into health care teams has been found to decrease costs,[5] determining which women may benefit from midwifery care could lead to cost-effective ways of improving maternal and infant health services.

This study investigates the ethnic differences associated with midwifery care. The objective was to document the association between maternal ethnicity and midwife attendance at birth from 1982 through 1989 and to determine how much of that association is explained by additional maternal characteristics. Although maternal characteristics associated with midwifery care have been reported,[2–4] differences between minority groups in the United States have not.

Methods

Natality files for 1982 through 1989, compiled by the National Center for Health Statistics, were used to examine US births attended by midwives.[12] To better compare physician- and midwife-attended births, this study population was limited to singleton births of infants weighing at least 2500 g. Births attended by someone other than a physician or midwife were excluded. Five categories of maternal ethnicity were defined by using the race and Hispanic origin data from the birth certificate.[12] Mothers of Hispanic origin were classified as Hispanic regardless of their race. In 1982, 23 states and the District of Columbia reported Hispanic origin of the mother; by 1989, 47 states reported Hispanic origin. However, even in the early 1980s, over 90% of Hispanic residents lived where Hispanic origin was recorded on the birth certificate.[13]

To control for the influence of additional maternal characteristics, the increased odds per year of midwife deliveries and the relative odds of midwifery use by ethnicity were estimated by using multiple logistic regression. Because the proportion of midwife-attended births is low, estimates from logistic regression correspond to the relative increase between years and the relative proportions among ethnic groups; however, these approximations slightly overstate trends for groups with more midwife utilization.[14] Regression models included maternal age, parity, education, first-trimester prenatal care, place of delivery (out of hospital vs in hospital), and marital status. Before 1989 maternal education was not recorded on birth certificates in California, Texas, and Washington; in 1988 and 1989, New York State (outside of New York City) did not collect information on

At the time of the study, Jennifer D. Parker was with the Division of Analysis, National Center for Health Statistics, Hyattsville, Md.

Requests for reprints should be sent to Jennifer D. Parker, PhD, 426 Warren Hall, School of Public Health, University of California, Berkeley, CA 94720.

This paper was accepted September 7, 1993.

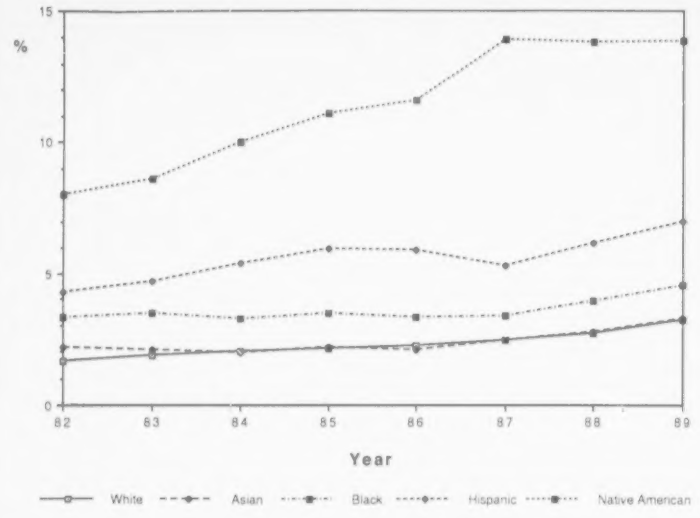


FIGURE 1—Percentage of US midwife-attended births of infants weighing at least 2500 g, by maternal ethnicity, 1982 through 1989.

TABLE 1—Average Percentage Increase and 95% Confidence Intervals (CIs) for Midwife Deliveries per Year, by Maternal Ethnicity, 1982 through 1989

	Unadjusted		Adjusted[a]	
Ethnicity	%	95% CI	%	95% CI
Non-Hispanic White	9.04	8.88, 9.19	7.73	7.57, 7.89
Black	4.06	2.83, 5.31	3.54	3.28, 3.80
Native American	9.77	8.34, 11.21	9.64	8.98, 10.30
Hispanic	6.74	5.46, 8.03	4.10	3.77, 4.43
Asian	7.77	6.26, 9.29	7.16	6.27, 8.06

[a]Adjusted for in hospital vs out of hospital, age, parity, education, first-trimester prenatal care, and marital status.

TABLE 2—Relative Odds for Midwife-Attended Delivery, by Maternal Ethnicity, 1982 through 1989

Ethnicity	AOR[a]	95% CI	AOR[b]	95% CI
Non-Hispanic White	Reference		Reference	
Black	1.59	1.58, 1.60	1.39	1.38, 1.40
Native American	5.54	5.47, 5.62	5.12	5.04, 5.19
Hispanic	2.56	2.54, 2.58	2.35	2.33, 2.37
Asian	1.09	1.07, 1.11	1.09	1.09, 1.14

Note. AOR = adjusted odds ratio; CI = confidence interval.
[a]Adjusted for year of delivery.
[b]Adjusted for year of delivery, in hospital vs out of hospital, age, parity, education, first-trimester prenatal care, and marital status.

education. This analysis was limited to the remaining states and the District of Columbia. Births missing information about maternal characteristics (fewer than 2% of the remaining births) were also excluded.

Results

Overall, 3.99% of singleton infants weighing more than 2500 g were delivered by a midwife in 1988, compared with 2.09% in 1982. An increase in midwife-attended births was apparent among all groups of women; however, the rate and pattern of increase differed by ethnicity (Figure 1). In 1989, 90% more White, 70% more Native American, 60% more Hispanic, 50% more Asian, and one third more Black women delivered with midwives than in 1982.

On average, midwifery care during this period increased over 7% per year among Native American, Asian, and White mothers, but increased less rapidly among Black and Hispanic mothers (Table 1). The impact of maternal characteristics varied by ethnicity. For example, after adjustment for demographic characteristics, the increased odds of being attended by a midwife rather than a physician dropped by almost one third among Hispanic mothers, but changed little among Native American mothers.

In spite of the rapid increase in midwife-attended births to White mothers during the 1980s, these mothers delivered with midwives less frequently than mothers of all other ethnic backgrounds (Table 2). Except for Asian mothers, adjustment for maternal factors reduced the ratio between White and non-White mothers by approximately 10%. However, Native American mothers remained approximately five times more likely and Hispanic mothers over twice as likely to be attended by a midwife than White mothers.

Discussion

This study documents ethnic differences in midwifery utilization throughout the 1980s in the United States. Compared with other mothers, for example, Native American mothers were the most likely, and White mothers were the least likely, to be attended by a midwife. However, among both groups the use of midwives increased rapidly during that time. Maternal factors, such as education and marital status, explained few of these differences; the unadjusted and adjusted estimates of midwife use were similar. The geographic distributions of ethnic groups, as well as of perinatal care providers, within the United States likely determined most of these differences.[3,4,8,9]

However, cultural preferences for physicians or midwives may also have

contributed to these findings. Acceptance of midwives may have expanded more rapidly among White mothers than among Black mothers, for example. Although White mothers were the least likely to be attended by a midwife, their use of midwives rose during the 1980s. This increase led to a rising proportion of midwife-attended White births during that time.[3]

Reports from 1989 birth certificates document maternal differences by type of midwife provider.[1,4] Unfortunately, before 1989, birth certificates were not that detailed; therefore, the term "midwife" in this study refers to both certified nurse-midwives and midwives with other training. Furthermore, Adams reported that more than 10% of certified nurse-midwives do not sign birth certificates,[15] also affecting comparisons between midwife and physician patients.

Nonetheless, the relatively high use of midwives by minority women and the overall increase in midwife-attended births throughout the 1980s suggest that midwifery is expanding to fill health care gaps. An understanding of ethnic differences in midwifery utilization may assist planners in effectively planning obstetric services in local communities. □

References

1. National Center for Health Statistics. Advance report of final natality statistics, 1989. *Month Vital Stat Rep.* December 1991;40(8)(suppl). DHHS publication PHS 92-1120.
2. National Center for Health Statistics, Taffel S. Midwife and out-of-hospital deliveries, United States. *Vital Health Stat S [21].* 1984; no. 40. DHHS publication PHS 84-1918.
3. Declercq ER. *The transformation of American midwifery:* 1975 to 1988. *Am J Public Health.* 1992;82:680–684.
4. Declercq ER. Where babies are born and who attends their births: findings from the revised 1989 United States Standard Certificate of Live Birth. *Obstet Gynecol.* 1993;81: 997–1004.
5. Bell KE, Mills JI. Certified nurse-midwife effectiveness in the health maintenance organization obstetric team. *Obstet Gynecol.* 1989;74:112–116.
6. Chambliss LR, Daly C, Medearis AL, Ames A, Kayne M, Paul R. The role of selection bias in comparing cesarean birth rates between physician and midwifery management. *Obstet Gynecol.* 1992;80:161–165.
7. Levy BS, Wilkinson FS, Marine WM. Reducing neonatal mortality rate with nurse-midwives. *Am J Obstet Gynecol.* 1971;109:50–58.
8. Slome C, Wetherbee H, Daly M, Christensen K, Meglen M, Thiede H. Effectiveness of certified nurse-midwives. *Am J Obstet Gynecol.* 1976;124:177–182.
9. Piechnik SL, Corbett MA. Reducing low birth weight among socioeconomically high-risk adolescent pregnancies. *J Nurse Midwifery.* 1985;30:88–98.
10. Knoll K. Certified nurse-midwives, nurse practitioners, and family practice physicians in the delivery of prenatal care. In: Merkatz IR, Thompson JE, eds. *New Perspectives on Prenatal Care.* New York, NY: Elsevier; 1990:611–623.
11. Mann PJ. San Francisco General Hospital nurse-midwifery practice: the first thousand births. *Am J Obstet Gynecol.* 1981;140: 676–682.
12. *Public Use Data Tape Documentation, United States Natality Files, 1978–1989.* Hyattsville, Md: National Center for Health Statistics; 1980–1991.
13. Ventura SJ. Births of Hispanic parentage, 1983 and 1984. *Monthly Vital Stat Rep.* July 24, 1987;36(4)(suppl 2). DHHS publication PHS 87-1120.
14. Hosmer DW, Lemenshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons; 1989.
15. Adams CJ. Nurse-midwifery practice in the United States, 1982 and 1987. *Am J Public Health.* 1989;79:1038–1039.

Exposure to Hair-Coloring Products and the Risk of Multiple Myeloma

Lisa J. Herrinton, PhD, Noel S. Weiss, MD, DrPH, Thomas D. Koepsell, MD, MPH, Janet R. Daling, PhD, John W. Taylor, MD, Joseph L. Lyon, MD, G. Marie Swanson, PhD, and Raymond S. Greenberg, MD, PhD

ABSTRACT

An interview study of persons diagnosed with multiple myeloma between 1977 and 1981 and suitable control subjects was conducted to test the hypothesis that exposure to hair dyes increases the risk of multiple myeloma. Among women, there was little evidence that prior regular use of hair dyes (odds ratio [OR] = 1.0; 95% confidence interval [CI] = 0.70, 1.4) or prior employment as a hairdresser (OR = 1.1; 95% CI = 0.43, 2.7) increased risk; however, the former comparison was limited by the lack of detailed information concerning the exposure. Among men, there was a modest association of regular use of hair dyes (OR = 1.5; 95% CI = 0.75, 2.9) with myeloma, but this was based on a small number of exposed persons. (*Am J Public Health.* 1994;84:1142–1144)

Introduction

We sought to investigate the hypothesis that exposure to hair dyes is associated with the risk of multiple myeloma by analyzing data from a population-based case-control study in which prior regular use of hair dyes and prior employment as a hairdresser were ascertained.

Methods

The data analyzed here were collected as part of a population-based case-control study of multiple myeloma that has been described previously.[1] Briefly, case patients under 82 years of age with incident multiple myeloma diagnosed from 1977 to 1981 were identified through cancer registries participating in the Surveillance, Epidemiology, and End-Results program in four geographic areas. Control subjects, matched so that their distributions of age, sex, and race were similar to the case patients, were identified using area sampling methods and random-digit dialing. A total of 689 case patients and 1681 control subjects, representing 89% and 83% of all eligible subjects, respectively, were identified.

Trained professional interviewers administered a standard questionnaire to 64% of the case patients and 99% of the control subjects. For the remaining subjects, who had either died or were ill, information was obtained from a spouse or relative (proxy respondent). Each subject's exposure status was determined up to a reference date, which was the date of diagnosis for the case patients and 1 year before the date of interview for the control subjects. Lifetime occupational histories were coded using the 1970 US Census codes, while prior regular use of hair dyes was determined from responses to the question: "Have you ever regularly used hair dyes?" Odds ratios (ORs) and exact 95% confidence intervals (CIs) were computed using logistic regression,[2] with adjustment for the matching variables of age (continuous), sex, race (White, other), and study center (Seattle, Utah [several counties], Detroit, Atlanta). Adjustment was also made for educational attainment (did not complete high school, high school graduate) because its inclusion in the logistic regression model influenced the odds ratios somewhat.

Results

Among self-responding women, a similar proportion of case patients and control subjects reported prior regular use of hair dyes (OR = 1.0; 95% CI = 0.70, 1.4) and prior employment as a hairdresser (OR = 1.1; 95% CI = 0.43, 2.7) (Table 1), and there was no consistent trend of increasing myeloma risk with increasing duration of employment (*P* trend, .558). Among self-responding men, there was a modest association of myeloma risk with reported prior use of hair dyes (OR = 1.5; 95% CI = 0.75, 2.9) and prior employment as a hairdresser (OR = 2.2; 95% CI = 0.18, 27), although both estimates of excess risk were based on small numbers of exposed subjects and were well within the limits of chance given no true association. There was little difference in the results when data from proxy respondents were included in the analysis.

At the time of the study, Lisa J. Herrinton, Noel S. Weiss, Thomas D. Koepsell, and Janet R. Daling were with the Department of Epidemiology, University of Washington, and the Division of Public Health Sciences, at the Fred Hutchinson Cancer Research Center, Seattle, Wash. Dr Herrinton is currently with Kaiser-Permanente in Oakland, Calif. John W. Taylor is with Group Health Cooperative of Puget Sound, Seattle. Joseph L. Lyon is with the School of Medicine, University of Utah, Salt Lake City. G. Marie Swanson is with Michigan State University's Cancer Center, East Lansing. Raymond S. Greenberg is with the School of Public Health, Emory University, Atlanta, Ga.

Requests for reprints should be sent to Noel S. Weiss, MD, DrPH, attention Carolyn Burns, Fred Hutchinson Cancer Research Center, 1124 Columbia MP-381, Seattle, WA 98104.

This paper was accepted October 21, 1993.

Discussion

An important limitation of these findings concerns the lack of precision in the question concerning personal use of hair dyes. A wide range of hair-coloring products may have elicited an affirmative response to the question concerning regular use of hair dyes, thereby obscuring a true exposure–disease relationship with a specific chemical. Also, some individuals may have incorrectly given a negative response to the question; such nondifferential misclassification also would have obscured any true association. In addition, our power to detect an association with a history of employment as a hairdresser was limited, as reflected by the width of the confidence intervals observed for these comparisons.

Experimental evidence that components in hair-coloring products cause cancer has been reviewed by Zahm and colleagues.[3] Briefly, the product formulations vary by manufacturer and over time. Darker formulations generally contain greater concentrations of dyes, while permanents contain oxidants that may affect the absorption of particular chemicals. Hairdressers may be more highly exposed to the chemicals in hair dye than are personal users of hair-coloring products, although an assessment of the relative exposures received by these two groups has not been conducted. Skin-painting experiments have given evidence for exposure through dermal absorption; it is also reasonable to believe that there is exposure through inhalation and inadvertent ingestion.

There have been four studies concerning the relationship between exposure to hair dyes and multiple myeloma risk. Two, using proportional measures of incidence[4] or mortality,[5] found evidence for an association with a history of employment as a cosmetologist or hairdresser (proportional incidence ratio, 467[4]; proportional mortality ratio, 619[5]). Follow-up of a cohort of 13 650 persons holding licenses for 5 years or longer prior to 1971 to practice cosmetology in Connecticut gave no evidence that myeloma risk was increased relative to that in the general Connecticut population (standardized incidence ratio, 0.69; 95% CI = 0.14, 2.0).[6]

However, in a population-based case–control study conducted in eastern Nebraska between 1983 and 1986, evidence was found that personal use of any hair-coloring products, and particularly permanents, was associated with myeloma risk (women: OR = 1.8; 95% CI = 0.9, 3.7; men: OR = 1.8; 95% CI = 0.5, 5.7), and there was a fairly consistent trend in risk with increasing duration of use of the products.[3] Two important differences between the Nebraska study and our own may account for the differences in our results. First, our ascertainment of personal use of hair-coloring products was limited and may have caused us to underestimate the association. As described previously, we did not obtain the names of the specific products used or information concerning color shade or permanence. The second difference concerns the sampling frames of the two studies: the Nebraska study included case patients diagnosed several years later than those in our study. It is possible that the time frame during which the Nebraska study was conducted, following reports that use of hair dyes may cause cancer, may have resulted in some degree of ascertainment bias. Prior regular use of hair dyes was associated not only with the occurrence of multiple myeloma in that study, but also with non-Hodgkin's lymphoma and Hodgkin's disease (although not with chronic lymphocytic leukemia). While it is possible that the etiologies of these three malignancies have some factors in common, it is also possible that the odds ratios for the associations of the individual malignancies with a history of using a hair-coloring product were all influenced by recall.

Our study provides little evidence that exposure to hair dyes increases myeloma risk. Nonetheless, because exposure to the chemicals present in hair-

TABLE 1—History of Use of Hair Dyes or of Prior Employment as a Hairdresser in Persons with Multiple Myeloma and in Control Subjects

Exposure	All Respondents[a]				Self-Respondents[b]			
	No. Case Patients	No. Control Subjects	OR[c]	95% CI	No. Case Patients	No. Control Subjects	OR[c]	95% CI
Women								
Ever regularly used hair dyes[c]								
No	205	477	1.0	. . .	151	471	1.0	. . .
Yes	114	267	1.1	0.83, 1.5	75	226	1.0	0.70, 1.4
Employed as a hairdresser for ≥ 6 mo								
Never	309	724	1.0	. . .	219	716	1.0	. . .
Ever	12	22	1.3	0.60, 2.7	7	21	1.1	0.43, 2.7
≤ 2 y	3	5	1.2	0.25, 5.3	2	5	0.94	0.15, 5.7
2–5 y	5	2	6.6	1.2, 36	2	2	5.7	0.78, 41
> 5 y	4	15	0.66	0.21, 2.0	3	14	0.69	0.19, 2.5
Men								
Ever regularly used hair dyes[c]								
No	343	892	1.0	. . .	228	882	1.0	. . .
Yes	17	39	1.3	0.70, 2.3	13	38	1.5	0.75, 2.9
Employed as a hairdresser for ≥ 6 mo								
Never	359	931	1.0	. . .	359	931	1.0	. . .
Ever	1	2	1.5	0.12, 17	1	2	2.2	0.18, 27

[a]Information missing for women, two control subjects and two case patients; men, two control subjects and one case patient.
[b]Information missing for women, two control subjects and one case patient.
[c]Odds ratio (OR) adjusted for age (continuous), race (White, other), study center (Seattle, Utah, Detroit, Atlanta), and educational attainment (did not complete high school, high school graduate). CI = confidence interval.

coloring products is widespread, it is reasonable to pursue this issue in future studies of hematologic malignancies, as well as in additional cohort studies of hairdressers. □

Acknowledgments

This research was supported by grants RO1 CA23350, R35 CA39779, and T32 CA09168 from the National Cancer Institute.

References

1. Koepsell TD, Daling JR, Weiss NS, et al. Antigenic stimulation and the occurrence of multiple myeloma. *Am J Epidemiol.* 1987; 126:1051–1062.
2. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods.* Belmont, Calif: Lifetime Learning Publications; 1982.
3. Zahm SH, Weisenburger DD, Babbitt PA, Saal RC, Vaught JB, Blair A. Use of hair coloring products and the risk of lymphoma, multiple myeloma, and chronic lymphocytic leukemia. *Am J Public Health.* 1992;82:990–997.
4. Guidotti S, Wright WE, Peters JM. Multiple myeloma in cosmetologists. *Am J Ind Med.* 1982;3:169–171.
5. Spinelli JJ, Gallagher RP, Band PR, Threlfall WJ. Multiple myeloma, leukemia, and cancer of the ovary in cosmetologists and hairdressers. *Am J Ind Med.* 1984;6:97–102.
6. Teta MJ, Walrath J, Meigs JW, Flannery JT. Cancer incidence among cosmetologists. *J Natl Cancer Inst.* 1984;72:1051–1057.

A Decision-Analytic Approach to Postexposure Rabies Prophylaxis

Scott B. Cantor, PhD, Richard D. Clover, MD, and Robert F. Thompson, MD

ABSTRACT

The risks and benefits of postexposure rabies prophylaxis were analyzed from clinical and economic perspectives. A decision-analytic model was constructed by using probability and outcome data from the literature and the state health department. Health outcomes were measured in quality-adjusted life years. In the base case (overweight adult male), treatment is optimal when the probability of animal rabidity is greater than 1 in 2000. Sensitivity analysis showed robustness in the treatment decision; however, the incremental cost-effectiveness ratio ($140 000/quality-adjusted life year) is sensitive to the rabidity probability. Treatment is optimal from the patient's perspective; however, it may not be cost-effective when the probability of rabidity is low. (*Am J Public Health.* 1994;84:1144–1148)

Introduction

Since the development of domestic animal control programs in the 1950s, the number of cases of human rabies in the United States has decreased from approximately 20 in 1952 to 0–5 per year today.[1] However, the decision to provide postexposure prophylaxis against rabies is still a frequent problem for public health departments, emergency rooms, and primary care physicians.[2] Appropriate management of persons possibly exposed to rabies depends on several factors, including the epidemiology of the disease, routes of transmission, and the perceived risk of infection. McCombie states that immunization decisions are affected not only by medical indications but also by factors such as reducing the patient or clinician's anxiety, preventing psychosomatic illness, reducing the fear of malpractice litigation, and fulfilling the clinician's need to intervene.[3] He suggests that formal medical decision analysis will lead to optimal treatment and more cost-effective decisions.

This article presents a decision analysis of the dilemma of whether or not to administer postexposure prophylaxis for rabies based on currently available medical information. The clinical case of an overweight 56-year-old White man who had been bitten by an opossum prompted this analysis.

Methods

A decision-analytic model was created to structure the decision whether or not to offer postexposure rabies prophylaxis.[4] As shown in Figure 1, treatment can be withheld or offered. If treatment is withheld and the animal was rabid, then the patient will die. If treatment is withheld and the animal was not rabid, normal life expectancy will follow.

Postexposure rabies immunization is assumed to be effective. However, adverse side effects can occur (Table 1). Minor reactions do not last longer than 1 month[5,6]; major reactions last no more than 3 months.[7–9]

A third strategy is plausible, in which the patient can be observed for 7 days to determine whether further treatment is necessary. However, the Centers for Disease Control and Prevention insist that treatment decisions about rabies must be made immediately.[10]

Probability of Rabidity

Once it has been established that an infection might have been transmitted

The authors are with the Department of Family Medicine, The University of Texas Medical Branch at Galveston, Galveston, Tex.

Requests for reprints should be sent to Scott B. Cantor, PhD, Department of Family Medicine, The University of Texas Medical Branch at Galveston, 301 University Blvd, Galveston, TX 77555-0853.

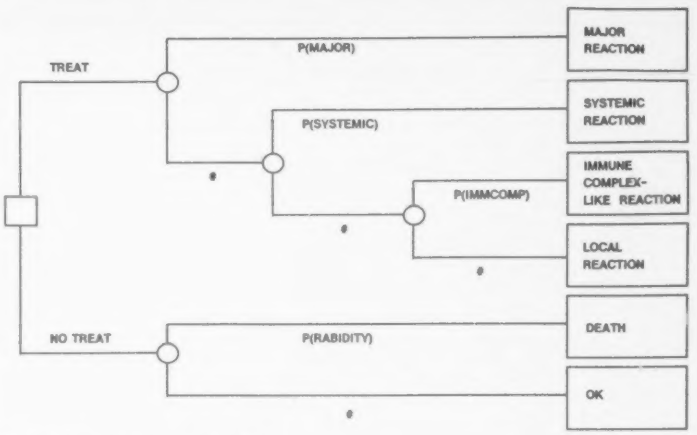
This paper was accepted November 3, 1993.

from animal to human through an open wound or bite, the likelihood of rabidity needs to be evaluated. The type of animal and the regional experience with rabies will determine the probability of rabidity. The municipal or state health department can provide data regarding the rabies status of animals recently tested.

However, an interesting but difficult methodological problem occurred in our estimation of the probability of rabidity when all previously examined animals were found not to be rabid (Texas Department of Health, unpublished data, 1992). In this case, the classical statistics approach of using the maximum likelihood estimate of zero did not provide a satisfactory answer. Decision analysis demands an explicit probability estimate; therefore, the alternative Bayesian approach of prior-to-posterior updating was chosen.[11]

In the Bayesian approach to estimating an unknown population parameter, the sample evidence is combined with a prior probability distribution, and the mean of the posterior probability distribution is computed for use as the parameter estimate. In this case, we needed to determine an initial prior distribution of the probability that a particular animal is rabid. The beta distribution is a good choice because of its ease of use and applicability to problems with an unknown population proportion. Specifically, if the prior probability distribution is beta, then the posterior probability distribution will also be beta with the mean equal to $(r + a)/(n + a + b)$, where a and b are the parameters of the beta distribution, n is the number of samples taken, and r is the number of rabid animals found in the sample.[12]

Without any previous guidance to estimate the probability of rabidity, we might have considered the beta distribution with parameters $a = 1$ and $b = 1$ (uniform distribution) as the prior distribution of the probability of rabidity. To err on the side of not treating, we chose the beta distribution with parameters $a = 1$ and $b = 101$ for the prior distribution. After we incorporated our sample of 752 nonrabid opossums, we computed that the posterior probability that the opossum in question was rabid was 1/854. This latter assumption was used in the base case analysis.



Note. # = complementary probability.

FIGURE 1—Decision-analytic model of the decision to administer postexposure rabies prophylaxis.

TABLE 1—Outcomes of the Postexposure Rabies Prophylaxis Decision

Outcome	Conditional Probability[a]	Duration of Adverse Reaction	Expected No. of Physician Visits	Vaccine	Complications
Treatment					
Local reaction (pain, erythema, swelling, or itching at injection site)	0.72	1 mo	1.5	Yes	None
Systemic reaction (headache, nausea, abdominal pain, muscle aches, dizziness)	0.22	1 mo	2.5	Yes	None
Immune complex–like reaction	0.08	1 mo	2.5	Yes	None
Major neurological reaction (resembling Guillain-Barré syndrome and other nervous system disorders)	0.000001	3 mo	4.0	Yes	14 hospital days
No treatment					
Animal rabid (death)			1.0	No	30 hospital days
Animal not rabid (normal survival)			1.0	No	None

Note. Physician office visit costs are considered in addition to hospital costs.
[a]See Figure 1.
Sources. Probability data are from references 7–10. The expected numbers of physician office visits are based on authors' estimates. Average length of stay data are based on reference 18.

Outcomes

The outcomes in the decision-analytic model are expressed in quality-adjusted life years.[13] In the case of an untreated bite from an animal that was not rabid, then the standard life expectancy is used[14]; for an untreated rabid bite, it is assumed that the quality-adjusted life expectancy is 0 quality-adjusted life years.

TABLE 2—Quality-Adjusted Life Expectancy Based on the Time Trade-Off Method: Outcomes after Postexposure Rabies Prophylaxis

Adverse Effect	Quality-Adjusted Life Expectancy
Local reaction	0.99 · (1/12) + (11/12) + (NORMLE-1)
Systemic reaction	uSYS · (1/12) + (11/12) + (NORMLE-1)
Immune complex–like reaction	uIMM · (1/12) + (11/12) + (NORMLE-1)
Major neurological reaction	uMAJ · (3/12) + (9/12) + (NORMLE-1)

Note. The amount of standard life expectancy is assigned the variable NORMLE. The variables uSYS, uIMM, and uMAJ represent the time trade-off utilities for systemic, immune complex–like, and major neurological reactions. The first two terms of each expression represent the quality-of-life adjustment for the first year of life after prophylaxis. The third term represents the remaining years of normal life expectancy.

TABLE 3—Quality-of-Life Adjustment Weights: Time Trade-Off Method of Utility Assessment

Adverse Effect	Mean[a]	SD	Median
Local reaction	0.99	. . .	. . .
Systemic reaction	0.56	0.19	6 mo (0.50 y)
Immune complex–like reaction	0.57	0.23	7 mo (0.58 y)
Major neurological reaction	0.31	0.20	3 mo (0.25 y)

Source. Survey of 15 University of Texas Department of Family Medicine faculty, residents, and students. Quality-of-life adjustment for local reaction was assumed to be .99.
[a]Fractional portion of 1 year in perfect health that is equivalent to 1 year in substandard health state.

TABLE 4—Results of Sensitivity Analysis for Base Case of 56-Year-Old White Male

Probability Variable	Threshold Probability
Rabidity	.0005
Systemic reaction	.5997
Immune complex–like reaction	.5378
Major neurological reaction	.0778
Time Trade-Off Parameter (Quality-of-Life Adjustment Weight)	**Threshold Value**
Local reaction	0.7803
Systemic reaction	Not found
Immune complex–like reaction	Not found
Major neurological reaction	Not found

We did quality-of-life utility assessment using the time trade-off method.[15] A sample of 15 faculty, residents, and medical students in a university hospital department of family medicine were asked, in questionnaire format, the number of months in perfect health they perceived to be equivalent to 12 months in a reduced health state. The mean responses were used to compute the quality-adjusted life expectancy for each outcome (Table 2).

Costs

Cost data were estimated from a national survey of hospital financial and operating statistics.[16] Data are based on costs, not charges, of all hospital services including overhead.[17] For all admissions, the national average cost per discharge was divided by the average length of stay to approximate an average cost of $722 per hospital day.

For every possible outcome modeled, the expected number of physician visits and length of hospital stay were estimated by a panel of academic family physicians. The more severe the adverse reaction, the greater the number of visits and the longer the hospital stay[18] (Table 1).

The cost of the vaccination is high and is proportional to the total body weight of the patient. For an overweight man weighing 135 kg, the vaccination series costs $606. For purposes of sensitivity analysis (see below), it was determined that the cost of vaccinations for a child weighing 30 kg would be $383 (Pasteur Mérieux, Lyons, France).

We analyzed the cost-effectiveness from a societal perspective using the outcome measure of dollars per quality-adjusted life year gained. We discounted costs and effectiveness at a 5% rate. In this model, there are no future costs; however, the future benefits of normal life expectancy were discounted to fairly compare the cost-effectiveness of postexposure rabies prophylaxis with other therapeutic interventions.[19]

Software and Sensitivity Analysis

Analysis was performed with SMLTREE 2.9 software.[20] Sensitivity analysis was performed to determine whether the optimal decision would be robust to changes in the value of the probability and utility parameters. Additional analyses were performed for a hypothetical 9-year-old Black female patient. The alternative scenario of a bitten child is a more frequent occurrence in primary care and provides a comparison.

Results

Table 3 summarizes the results of the time trade-off survey. Surprisingly, the immune complex-like reaction was perceived as nearly equivalent to the minor systemic reaction. However, the neurologic reaction was rated worse than either minor reaction, consistent with what was anticipated.

We then used the survey results to determine the optimal strategy, using the criterion of expected quality-adjusted life-year maximization. The treatment strategy yielded 19.989 expected quality-adjusted life years; the no-treatment strategy yielded 19.977 expected quality-adjusted life years. Thus, treatment is preferred to no treatment by approximately 4.5 quality-adjusted life days.

One-way sensitivity analysis revealed that treatment would be the optimal decision when the probability of animal rabidity was greater than 1 in 2000 (Table 4). Threshold values for the other probabilities of adverse reaction were also found to be much greater than the base case parameters.

Changes in individual utility parameters also had no effect on the optimal decision, with the exception of the time trade-off parameter for local reaction. If 1 year of having an adverse local reaction were considered equivalent to less than 9 months of perfect health, then no treat-

ment would be the optimal decision. Thus, when quality-adjusted life years are used as the outcome measure, treatment is a robust optimal strategy.

An economic analysis of the treatment decision can be based on the incremental cost-effectiveness ratio of treatment compared with no treatment, defined as the difference in costs divided by the change in quality-adjusted life expectancy. As shown above, it is beneficial from the patient's perspective to treat; however, treatment is costly. On an expected cost basis, the treatment strategy is more expensive than the no-treatment strategy ($721 vs $90). From the societal perspective, the incremental cost-effectiveness ratio was approximately $140 000 per additional quality-adjusted life year gained. In comparison, previously analyzed interventions such as treatment of moderate hypertension and screening for osteoporosis have been shown to be cost-effective at a cost of $25 000 to $40 000 per quality-adjusted life year gained.[21,22]

However, the cost-effectiveness ratio is very sensitive to the probability of rabidity. Table 5 shows the variability of the cost-effectiveness ratio for many animals. For high probabilities of rabidity, treatment is a dominant strategy, saving more lives at a lower expected cost by avoiding the likely expensive hospitalization of a dying patient. As the probability of rabidity decreases, the cost-effectiveness ratio increases exponentially. For low probabilities of rabidity, no treatment is a dominant strategy; treatment is more expensive than withholding treatment with no increase in life expectancy.

When the same analyses were performed for a 9-year-old Black female, the results changed—but not dramatically. In this case, the threshold probability of rabidity was 1 in 5000. Less than 1 quality-adjusted life month was expected to be gained by treatment; however, the expected benefits from treatment cost less than $32 500 per quality-adjusted life year gained. The health benefits and cost-effectiveness of treatment depend not only on the probability of rabidity, but also on the patient's age and body weight.

TABLE 5—Sensitivity Analysis of Cost-Effectiveness Ratio Based on Probability of Animal Rabidity

Probability of Rabidity[a]	Animal	Incremental Cost-Effectiveness Ratio
1 in 3	Skunk	TREAT dominating strategy[b]
1 in 10	Bat	TREAT dominating strategy[b]
1 in 100	Dog (wild)	$3 662 per QALY gained
1 in 200	Cat	$10 028 per QALY gained
1 in 854	Opossum	$139 988 per QALY gained
1 in 1000		$278 144 per QALY gained
1 in 10 000		NO TREAT dominating strategy[c]
1 in 100 000		NO TREAT dominating strategy[c]

Note. QALY = quality-adjusted life year.
[a]Probability of rabidity is based on unpublished data from the Texas Department of Health, 1987–1991.
[b]TREAT is the dominating strategy because costs are lower and expected QALYs are higher than those associated with the NO TREAT strategy (based on high probability of rabidity).
[c]NO TREAT is the dominating strategy because costs are lower and expected QALYs are higher than those associated with the TREAT strategy (based on low probability of rabidity).

Discussion

A decision-analytic model of postexposure rabies prophylaxis has not been previously published. Because the vaccine has potentially serious, albeit rare, adverse effects, it is erroneous to assume that vaccinations should always be administered. Thus, it is important to approximate the probability of rabidity. Formal decision analysis is an important, appropriate methodology to guide clinical management.

For many animal species, the probability of a particular animal's being rabid is very slight. If physicians can reduce the tendency to overestimate the probabilities of unlikely outcomes,[23] then the bias to overtreat may be reduced. With better understanding of small probabilities, clinicians can make better decisions.

The small differences in quality-adjusted life expectancy for the two strategies may lead some to the conclusion that the postexposure rabies prophylaxis decision is a "toss up."[24] Kassirer and Pauker suggest that exogenous factors not considered in the original model should then be included. In this case, patient risk aversion,[25] patient anxiety,[26,27] and physician preferences[28] would all support the decision to treat.

This case exemplifies a classic dilemma of medical decision making: what may be optimal treatment for the patient may not be cost-effective for society. In the case of the decision to treat after potential rabies exposure, the decision-analytic model showed that treatment was indicated, even though the gain in quality-adjusted life expectancy was less than 1 month. As was shown in the threshold analysis, treatment was indicated for probability of rabidity greater than 1 in 2000 for the adult case and 1 in 5000 for the child case. However, based on an incremental cost-effectiveness analysis, the monetary amount needed to achieve that benefit is significant, although it may be barely consistent with other therapeutic interventions considered to be cost-effective. □

Acknowledgments

This research was funded in part by a grant from the Bureau of Health Professions, Health Resources and Services Administration (D32-PE16033).

The authors thank Dr Melchor Cardenas for research assistance, Mr David Connaughton and Ms Nancy Maguire for providing relevant cost data, Dr William Crump for his valuable comments and suggestions, and Ms Dee Ann Forasiepi for preparing the manuscript.

This paper was presented at the 15th Annual Meeting of the Society for Medical Decision Making, Research Triangle Park, NC, October 1993.

References

1. Reid-Sanden FL, Dobbins JG, Smith JS, Fishbein DB. Rabies surveillance, United States during 1989. *J Am Vet Med Assoc.* 1990;197:1571–1583.
2. Fishbein DB, Arcangeli S. Rabies prevention in primary care: a four-step approach. *Postgrad Med.* 1987;82:83–95.
3. McCombie SC. The politics of immunization in public health. *Soc Sci Med.* 1989;28:843–849.
4. Kassirer JP, Moskowitz AJ, Lau J, Pauker SG. Decision analysis: A progress report. *Ann Intern Med.* 1987;106:275–291.
5. Centers for Disease Control. Systemic allergic reactions following immunization with human diploid cell rabies vaccine. *MMWR.* 1984;33:185–187.
6. Dreesen DW, Bernard KW, Parker RA, Deutsch AJ, Brown J. Immune complex-like disease in 23 persons following a booster dose of rabies human diploid cell vaccine. *Vaccine.* 1986;4:45–49.
7. Boe E, Nyland H. Guillain-Barré syndrome after vaccination with human dip-

loid cell rabies vaccine. *Scand J Infect Dis.* 1980;12:231–232.
8. Bernard KW, Smith PW, Kader FJ, Moran MJ. Neuroparalytic illness and human diploid cell rabies vaccine. *JAMA.* 1982;248: 3136–3138.
9. Knittel T, Ramadori G, Mayet WJ, Löhr H, Meyer zum Büschenfelde KH. Guillain-Barré syndrome and human diploid cell rabies vaccine. *Lancet.* 1989;1:1334–1335.
10. Centers for Disease Control. Rabies prevention—United States, 1991: recommendations of the Immunization Practices Advisory Committee (ACIP). *MMWR.* 1991;40(RR-3):1–19.
11. Pratt JW, Raiffa H, Schlaifer R. *Introduction to Statistical Decision Theory.* New York, NY: McGraw-Hill; 1965.
12. Chew V. Point estimation of the parameter of the binomial distribution. *Am Statistician.* 1971;25:47–50.
13. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Operations Res.* 1980;28:206–224.
14. Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical Decision Making.* Boston, Mass: Butterworth; 1988: appendix 2—life table, 373–378.
15. McNeil BJ, Weischelbaum R, Pauker SG. Speech and survival. Tradeoffs between quality and quantity of life in laryngeal cancer. *N Engl J Med.* 1981;305:982–987.
16. Cleverly WO. *The 1992 Almanac of Hospital Financial and Operating Indicators.* Columbus, Ohio: The Center for Healthcare Industry Performance Studies; 1992.
17. Weinstein MC. Economic assessments of medical practices and technologies. *Med Decis Making.* 1981;1:309–330.
18. Health Care Investment Analysis, Inc, and Ernst and Young. *The Medicare DRG Handbook: Comparative Clinical and Financial Standards.* Baltimore, Md: Health Care Investment Analysis, Inc; 1992.
19. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med.* 1977;296: 716–721.
20. Hollenberg J. *SMLTREE 2.9.* New York, NY: Dept of Medicine, Cornell University Medical College; 1989.
21. Tosteson ANA, Rosenthal DI, Melton LJ, Weinstein MC. Cost effectiveness of screening perimenopausal white women for osteoporosis: bone densitometry and hormone replacement therapy. *Ann Intern Med.* 1990;113:594–603.
22. Weinstein MC, Stason WB. Cost-effectiveness of interventions to prevent or treat coronary heart disease. *Annu Rev Public Health.* 1989;79:1556–1558.
23. Kahneman D, Tversky A. The psychology of preferences. *Sci Am.* 1982;246:160–173.
24. Kassirer JP, Pauker SG. The toss up. *N Engl J Med.* 1981;305:1467–1469.
25. McNeil BJ, Weischelbaum R, Pauker SG. Fallacy of the five-year survival in lung cancer. *N Engl J Med.* 1978;299:1397–1401.
26. Asch DA, Patton JP, Hershey JC. Knowing for the sake of knowing. *Med Decis Making.* 1990;10:47–57.
27. Cantor SB. *A Decision Analytic Approach to Prenatal Diagnosis.* Cambridge, Mass: Harvard University; 1991. Unpublished doctoral dissertation.
28. Brett AS. Hidden ethical issues in clinical decision analysis. *N Engl J Med.* 1981;305: 1150–1152.

Mass Media and School Interventions for Cigarette Smoking Prevention: Effects 2 Years after Completion

ABSTRACT

The long-term cigarette smoking prevention effects of mass media and school interventions were assessed. Adolescents in two communities received both mass media and school interventions; those in two matching communities received only school interventions. Surveys of 5458 students were conducted at baseline in grades 4 through 6 and 2 years after the 4-year interventions were completed, when students were in grades 10 through 12. Students exposed to the media-plus-school interventions were found to be at lower risk for weekly smoking (odds ratio = 0.62, 95% confidence interval = 0.49, 0.78) than those receiving school interventions only, indicating that the effects of the combined interventions persisted 2 years after the interventions' completion. (*Am J Public Health.* 1994;84:1148–1150)

Brian S. Flynn, ScD, John K. Worden, PhD, Roger H. Secker-Walker, MB, BChir, Phyllis L. Pirie, PhD, Gary J. Badger, MS, Joseph H. Carpenter, MA, and Berta M. Geller, EdD

Introduction

We previously reported on significant reductions in adolescent cigarette smoking achieved by mass media interventions combined with school programs.[1,2] Since other smoking prevention studies have shown that effects were lost over time, we conducted additional surveys 2 years after these interventions ended.[3,4]

Methods

Baseline surveys were administered to 5458 students in grades 4 through 6 from four communities in the United States. Those in two communities received mass media and school interventions for 4 years; those in two matched communities received only the school interventions. Interventions were focused on cigarette smoking prevention.[1,2] When these students were in grades 10 to 12, we attempted to follow up all cohort members.

Study Samples

Within the four study communities, equivalence of treatment groups was increased by selecting schools in census tracts having characteristics indicating higher risk for cigarette smoking. In media-plus-school tracts, 68% of adults

Brian S. Flynn, John K. Worden, Roger H. Secker-Walker, and Berta M. Geller are with the Office of Health Promotion Research; additionally, Drs Flynn, Worden, and Geller are with the Department of Family Practice; and Drs Flynn, Worden, and Secker-Walker are with the Vermont Cancer Center, College of Medicine, University of Vermont, Burlington. Gary J. Badger and Joseph H. Carpenter are with the Medical Biostatistics Unit, College of Medicine, University of Vermont. Phyllis L. Pirie is with the Division of Epidemiology, School of Public Health, University of Minnesota, Minneapolis.

Requests for reprints should be sent to Brian S. Flynn, ScD, Office of Health Promotion Research, College of Medicine, University of Vermont, One South Prospect, #1308, Burlington, VT 05401.

This paper was accepted March 8, 1994.

were high school graduates, and median household income was $14 186. In school-program-only tracts, 74% were graduates and household income was $14 994.

The full-exposure sample consisted of cohort members participating in all six school surveys; this sample was assumed to have received the complete intervention. The complete-follow-up sample included all cohort members assessed in grades 10 to 12 regardless of survey participation during intervening years; these students received varying intervention exposure because some moved away or left school after the baseline survey.

Data Collection and Measurement

All school surveys were conducted by the University of Vermont. For cohort members who were not present for the follow-up school survey, telephone interviews were conducted by the University of Minnesota.[5] The primary outcome measure assessed cigarettes smoked in the past week; responses greater than zero were classified as weekly smokers. Two additional smoking measures, daily smoking and self-selected smoking category, were included to examine consistency of results.[6,7] Saliva samples were obtained during school surveys as a bogus pipeline procedure to encourage more accurate self-reports.[8] Self-reports of alcohol use were measured to assess substance use behaviors not targeted by the interventions.

Statistical Analyses

The study design was based on the individual as the unit of analysis. For analyses based on this perspective, stepwise logistic regression was used to assess treatment effects while adjusting for potential confounding variables. For analyses concerned with the complete-follow-up sample, additional predictors were included for type of follow-up survey and years of intervention exposure. Inclusion of variables was based on significance of the F-to-enter ($P < .05$) and improvement in model fit (Hosmer-Lemeshow test).

Consistency between units of allocation to treatment and units of analysis has been emphasized in the recent program evaluation literature.[9] Because study communities were assigned to treatment groups, analyses of variance based on this perspective are presented. These analyses have low statistical power because of the small number of community units available. Community was considered a random factor nested within treatment; school, a random factor nested within community; and individual, a random factor nested within school. Significance associated with treatment differences was determined using community within treatment as the error term.[10]

TABLE 1—Follow-Up Survey Participation Status of Baseline Cohort

	No. (n = 5458)	%
Follow-up successful	4670	86
School survey	2863	
Telephone survey	1807	
Follow-up not successful	788	14
Refused in school	86	
Refused over telephone	63	
Incomplete telephone interview	4	
Not reached by telephone	122	
Not located by telephone	238	
School had no tracking data	237	
Excluded by school (special education students)	29	
Deceased	9	

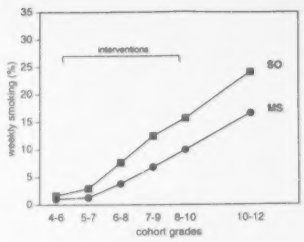


FIGURE 1—Weekly smoking among those participating in all six school surveys (full-exposure sample) in the media-plus-school (MS) and the school-program-only (SO) treatment groups (n = 2086).

Results

Of the 5458 students participating in baseline surveys, 38% (n = 2086) participated in the five subsequent school surveys and constituted the full-exposure sample; 86% (n = 4670) participated in surveys of grades 10 to 12 and constituted the complete-follow-up sample (Table 1). In the full-exposure sample, the media-plus-school group had a larger proportion of females, a group at higher risk for becoming smokers, than the school-program-only group. In the complete-follow-up sample, the media-plus-school group also were more likely to report sibling smokers.

Changes in weekly smoking for the full-exposure sample are shown in Figure 1. When 1991 weekly smoking was regressed on potential explanatory variables, five predictors met the inclusion criteria (Table 2). The odds ratio for being a smoker in the media-plus-school group was 0.62, indicating reduced risk. Similar models were obtained for the alternative smoking measures. Analyses of variance based on community as the analytic unit showed significant differences between treatment groups both before adjusting for covariates ($F_{1,2} = 24.6$; $P = .04$) and after adjustments ($F_{1,2} = 92.0$; $P = .01$).

When 1991 weekly smoking was regressed on potential explanatory variables in the complete-follow-up sample, five predictors again met the inclusion criteria (Table 2). In this model, the odds ratio for being a smoker in the media-plus-school group was 0.79. For analyses based on community as the unit, this difference was not statistically significant ($F_{1,2} = 4.38$; $P = .17$).

Self-reports of alcohol use were examined using similar logistic regression models. No treatment effects were observed.

Discussion

These results extend earlier findings of the smoking prevention efficacy of media and school interventions in two ways: (1) effects persisted 2 years after interventions ended; and (2) effects were observed among a more diverse sample of the original cohort.

The treatment groups had reasonably equivalent baseline characteristics but tended to show the media-plus-school group to be at higher risk for smoking; census tracts from which this group were drawn also showed higher population risks for smoking. Intervention effects were specific to cigarette smoking and did

TABLE 2—Logistic Regression Results Assessing Risk of Smoking among Cohort Members

Predictor	Outcome Measures: Weekly Smoking — Entry Order	Odds Ratio	95% Confidence Interval	Smoking Category — Entry Order	Odds Ratio	95% Confidence Interval	Daily Smoking — Entry Order	Odds Ratio	95% Confidence Interval
Full-exposure sample									
	(n = 1939; *P* = .77)[a]			(n = 1988; *P* = .40)			(n = 1925; *P* = .99)		
Media-plus-school	1	0.62	0.49, 0.78	2	0.61	0.47, 0.77	2	0.60	0.46, 0.78
Baseline smoker	2	3.38	1.47, 7.74	3	3.20	1.40, 7.31	4	2.96	1.20, 7.30
Female gender	3	1.49	1.18, 1.87	4	1.46	1.14, 1.87	5	1.36	1.04, 1.79
Parental smoking	4	1.39	1.10, 1.75	1	1.56	1.21, 2.00	1	1.80	1.35, 2.35
Sibling smoking	5	1.50	1.01, 2.22	5	1.75	1.18, 2.59	3	1.86	1.22, 2.84
Complete-follow-up sample									
	(n = 4307; *P* = .76)			(n = 4392; *P* = .60)			(n = 4304; *P* = .77)		
Telephone survey	1	2.02	1.76, 2.32	1	2.34	2.03, 2.69	1	2.40	2.06, 2.78
Baseline smoker	2	3.29	2.24, 4.84	3	3.45	2.35, 5.08	3	3.11	2.11, 4.58
Parental smoking	3	1.51	1.30, 1.74	2	1.73	1.49, 2.01	2	1.91	1.63, 2.25
Media-plus-school	4	0.79	0.69, 0.91	5	0.78	0.68, 0.90	4	0.78	0.67, 0.90
Sibling smoking	5	1.39	1.12, 1.73	4	1.50	1.21, 1.86	5	1.46	1.16, 1.83

[a]*P* values represent significance of Hosmer-Lemeshow goodness-of-fit test.

not reflect a general proneness toward less substance use in the media-plus-school group. These findings confirm that sample differences were unlikely explanations for observed effects.[1] Confidence in the findings was increased by the consistency of results across different smoking measures.

For the full-exposure sample, both individual and community-based analytic approaches provided evidence of significant treatment effects. For the complete-follow-up sample, the primary analysis showed significant effects while the more conservative approach did not. The strength and specificity of the effects, their consistency, and their agreement with earlier results, however, support a conclusion that treatment differences were observed in the latter sample.

These results should be generalizable despite demographic differences between study areas and the US population. Census data indicate that the cohort represented a lower-income sector of the US adolescent population. High levels of cohort recruitment and retention provided an unusually complete picture of long-term effects within this group.

Other studies have evaluated multifaceted smoking prevention approaches.[7,11–14] This group of studies suggests that school smoking prevention programs can have substantial effects when supported by efforts outside of schools to influence perceptions of peer and community norms. The present study shows that sustained smoking prevention effects can be achieved efficiently through schools and mass media, two of the most powerful influences on young people. □

Acknowledgments

This study was supported by research grants CA38395 and CA22435 awarded by the National Cancer Institute, and by research grant HL17292 awarded by the National Heart, Lung, and Blood Institute.

We are grateful to Mary Noonan and Dorothy Myer for administering the school surveys, and to Karen Virnig for her leadership of the telephone interview effort.

References

1. Flynn BS, Worden JK, Secker-Walker RH, Badger GJ, Geller BM, Costanza MC. Prevention of cigarette smoking through mass media intervention and school programs. *Am J Public Health.* 1992;82:827–834.
2. Worden JK, Flynn BS, Geller BM, et al. Development of a smoking prevention mass media program using diagnostic and formative research. *Prev Med.* 1988;17:531–558.
3. Murray D, Davis-Hearn M, Goldman A. Four- and five-year follow-up from four seventh grade smoking prevention strategies. *J Behav Med.* 1988;11:394–405.
4. Flay B, Koepke D, Thomson S. Six-year follow-up of the first Waterloo school smoking prevention trial. *Am J Public Health.* 1989;79:1371–1376.
5. Pirie PL, Murray DM, Luepker RV. Smoking prevalence in a cohort of adolescents, including absentees, dropouts, and transfers. *Am J Public Health.* 1988;78:176–178.
6. Pechacek T, Murray D, Luepker R, Mittelmark M, Johnson C, Shutz J. Measurement of adolescent smoking behavior: rationale and methods. *J Behav Med.* 1984;7:123–140.
7. Perry CL, Kelder SH, Murray DM, Klepp K-I. Communitywide smoking prevention: long-term outcomes of the Minnesota Heart Health Program and the Class of 1989 Study. *Am J Public Health.* 1992;82:1210–1216.
8. Murray DM, Perry C. The measurement of substance use among adolescents: when is the bogus pipeline method needed? *Addict Behav.* 1987;12:225–233.
9. Murray DM, Hannan PJ. Planning for the appropriate analysis in school-based drug-use prevention studies. *J Consult Clin Psychol.* 1990;58:458–468.
10. *SAS User's Guide, Statistics Version 6.03.* Cary, NC: SAS Institute; 1988.
11. Bauman K, Padgett C, Koch G. A media-based campaign to encourage personal communication among adolescents about not smoking cigarettes: participation, selection, and consequences. *Health Educ Res.* 1989;4:35–44.
12. Bauman KE, LaPrelle J, Brown JD, Koch GG, Padgett CA. Influence of three mass media campaigns on variables related to adolescent smoking: results of a field experiment. *Am J Public Health.* 1991;81:597–604.
13. Pierce J. Time to ban advertising and continue the Quit For Life campaigns. *Med J Aust.* 1990;152:113.
14. Pentz M, Dwyer J, MacKinnon D, et al. A multicommunity trial for primary prevention of adolescent drug abuse: effects on drug use prevalence. *JAMA.* 1989;261:3259–3266.

Counseling Adolescents for Smoking Prevention: A Survey of Primary Care Physicians and Dentists

David I. Gregorio, PhD, MS

ABSTRACT

Pediatricians, family practitioners, and dentists were questioned in the spring of 1992 about whether they counseled adolescent patients (10 through 18 years old) not to smoke. Of the 674 questionnaires distributed, 443 (66%) were returned and analyzed. Most practitioners could not estimate cigarette use among their adolescent patients. Prevention counseling occurred infrequently, and least often among younger patients. Physicians were more likely than dentists to counsel adolescents. The data highlight the need for continuing training of primary health care practitioners about the importance of counseling adolescents not to smoke. (*Am J Public Health.* 1994;84:1151–1153)

Introduction

After 3 decades of intense warnings about the hazards of tobacco use, smoking initiation rates among our nation's youth remain alarmingly high. Currently, one in six of our nation's children may become regular smokers, and one quarter of them, or 5 million individuals, will die unnecessarily from their addiction.[1] Prevention of smoking initiation is an essential strategy to lessen the number of persons addicted to nicotine and avoid the health effects of that habit.

It has long been appreciated that primary health care providers can play significant roles in helping young patients to avoid the use of tobacco.[2,3] They are knowledgeable about the hazards of smoking, they often serve as powerful role models,[4] and they are likely to have frequent contact with these patients.[5] Recently issued Guidelines for Adolescent Preventive Services recommend that persons between the ages of 11 and 21 years be screened annually for cigarette use and counseled to avoid tobacco and other abusable substances.[6]

Because the role that practitioners play in smoking prevention is not fully understood,[7] findings from a cross-sectional survey of Connecticut pediatricians, family practitioners, and general dentists are reported here.

Methods

Early in 1992, pediatricians (n = 156), family practitioners (n = 261), and general dentists (n = 257) from across the state were sent a two-page questionnaire on smoking prevention counseling of adolescent patients (10 through 18 years old). Random samples, accounting for roughly one third of practicing pediatricians and general dentists and two thirds of family physicians, were drawn from membership rosters provided by the respective state associations.

The initial mailing of 674 questionnaires yielded 354 (52%) responses. Six weeks later a follow-up request was sent to the 320 nonrespondents, producing an additional 89 replies for a total response rate of 66%. Pediatricians were somewhat more responsive to the initial mailing than either family practitioners or dentists, but overall responses were equivalent across groups (67%, 66%, and 64%, respectively). Initial and follow-up replies did not differ noticeably and were combined for analysis. Unfortunately, the characteristics of responders, nonresponders, and nonmembers of professional organizations could not be compared.

As a group, respondents were predominantly male (86%), never smokers (66%), and solo practitioners (55%) with substantial patient care experience (median experience was 14 years). Seven respondents (1.6%) reported being current smokers. Dentists, compared with physicians, were more likely to be solo practitioners (62% vs 51%, respectively) and more likely to have practiced more than 20 years (36% vs 23%, respectively). They were as likely as physicians to be male and to have a history of cigarette use.

Respondents were asked to estimate the proportions of their patients aged 10 through 12, 13 through 15, and 16 through 18 years who smoked. They were also asked to use selected descriptors ("always," "frequently," "occasionally," "seldom or never") to characterize how often they counseled young patients to avoid smoking and to indicate why they did not always counsel young patients. Counseling was not specifically defined and could have been interpreted by respondents as anything from a brief

The author is with the Department of Community Medicine and Health Care, University of Connecticut School of Medicine, Farmington, Conn.

Requests for reprints should be sent to David I. Gregorio, PhD, MS, Department of Community Medicine and Health Care, University of Connecticut School of Medicine, Farmington, CT 06030-1910.

This paper was accepted November 24, 1994.

Editor's Note. See related editorial by Shelov (p 1066) in this issue.

TABLE 1—Reported Frequency of Smoking Prevention Counseling, by Patient Age Group and Practitioner Specialty

Frequency of Counseling	Pediatricians (n = 105)		Family Physicians (n = 172)		General Dentists (n = 166)		Total (n = 443)	
	No.	%	No.	%	No.	%	No.	%
Patients aged 10–12 y								
Seldom or never	24	23	39	25	95	65	158	39
Occasionally	26	25	38	24	30	21	94	23
Frequently	33	32	42	27	9	6	84	21
Always	21	20	38	24	12	8	71	17
Not reported	1	...	15	...	20	...	36	...
$\chi^2 = 76.7; P < .01$								
Patients aged 13–15 y								
Seldom or never	5	5	15	9	72	47	92	22
Occasionally	15	15	33	20	55	37	103	24
Frequently	43	42	53	32	15	10	111	26
Always	40	39	65	39	12	8	117	28
Not reported	2	...	6	...	12	...	20	...
$\chi^2 = 145.7; P < .01$								
Patients aged 16–18 y								
Seldom or never	1	1	3	2	50	31	54	12
Occasionally	9	9	19	11	72	44	100	23
Frequently	43	42	61	36	26	16	130	30
Always	50	48	88	51	15	9	153	35
Not reported	2	...	1	...	3	...	6	...
$\chi^2 = 189.3; P < .01$								

Note. The percentages shown are percentages of respondents who answered the question about counseling patients in the indicated age group, not percentages of the total number of respondents.

TABLE 2—Perceived Barriers to Routine Counseling of Adolescents to Avoid Tobacco Use, by Practitioner Specialty

	No. Responding			
	Pediatricians (n = 105)	Family Physicians (n = 172)	General Dentists (n = 166)	Total (n = 443)
Counseling is generally ineffective in altering behavior of adolescents	6	15	27	48
Counseling is inconvenient in my practice setting	11	9	23	43
I don't feel qualified to counsel patients on this matter	0	0	21	21
It's difficult to know which of my patients are truly at risk	6	21	51	78
Compensation for wellness care is inadequate	3	8	7	18
Total no. (%) perceiving barriers*	26 (25%)	42 (24%)	90 (54%)	158 (36%)

*The number perceiving barriers does not equal the number of barriers reported because respondents could cite more than one barrier when answering the question. The percentage refers to the proportion of practitioners within a given specialty who perceived barrier(s) to counseling.

warning about tobacco use to more deliberate counseling sessions. Lastly, respondents were asked if they recorded tobacco use histories and/or episodes of tobacco counseling in their patients' medical records.

Results

Perceived Smoking Prevalence

Relatively few practitioners were familiar with cigarette use among their adolescent patients. Only one half of family practitioners (48%), few pediatricians (38%), and fewer dentists (26%) could estimate smoking prevalence within these age groups. Moreover, those making estimates tended to underestimate cigarette use; one half of respondents estimated the prevalence among 13- through 15-year-old patients to be less than 5% and one third estimated the prevalence among 16- through 18-year-olds to be less than 10%. (National estimates are that 8% to 11% of 13- through 15-year-olds and 16% of 16- through 18-year-olds use cigarettes frequently.[8])

Smoking Prevention Counseling Practices

Perhaps because practitioners do not perceive many adolescents as at risk, few counsel them about tobacco use (Table 1). Only 35% of the respondents reported "always" counseling 16- through 18-year-old patients, although another 30% indicated that counseling occurred "frequently." One half of the physicians, but fewer than 1 in 10 dentists (9%), reported "always" counseling this age group. Almost one third (31%) of dentists "seldom or never" counseled 16- through 18-year-olds about smoking.

Counseling decreased across all specialties for younger patients. One in five respondents (22%) "seldom or never" counseled 13- through 15-year-olds. Dentists were more likely than physicians to give this response (47% vs 7%, respectively). Only one quarter of the respondents (28%) indicated that they "always" counseled 13- through 15-year-olds, although a similar proportion of respondents described their counseling of this group as "frequent."

Only 38% of the respondents indicated that their 10- through 12-year-old patients were "frequently" counseled. Less than 20% of the respondents reported "always" counseling patients in this age group. Counseling by dentists was considerably below the level reported by physicians (i.e., those who reported "frequently" or "always" counseling patients in this age group were 14% and 48% of respondents, respectively).

Counseling practices of physicians and dentists were reexamined by sepa-

rately controlling for experience (≤ 14 years vs > 14 years), practice arrangement (solo vs group), smoking history (ever vs never), and sex. The patterns reported here were not affected by these procedures.

Barriers to Routine Counseling

Eighty-one respondents (51% of those who cited barriers to regular counseling) believed that "counseling is generally ineffective in altering behavior of adolescents"; an additional 78 respondents (49% of those who cited barriers) answered, "It's difficult to know which of my patients are truly at risk" (Table 2). Twenty-one dentists indicated "I don't feel qualified to counsel patients on this matter."

Record Keeping of Prevention Counseling

Most physicians (66%) routinely noted tobacco use by their adolescent patients in the patients' medical records, and many (45%) recorded counseling episodes as well. Dentists tended to record neither smoking histories (37% reported doing so) nor counseling (19%) in patients' records. (Both differences between physicians and dentists are significant at $P < .01$).

Discussion

Primary care providers are uniquely situated to help adolescents avoid experimentation with cigarettes. Providing information on the health effects of tobacco use along with incentives and skills to resist pressures to use tobacco can prevent some adolescents from becoming smokers.[9,10] Current practice guidelines recommend that prevention counseling be a regular part of health care for this population.[6] Unfortunately, active intervention by practitioners with adolescent patients is alarmingly low. These findings, which may be "best case" estimates because of a tendency among respondents to offer "socially approving" replies, indicate that counseling, especially for the youngest adolescents (10 through 12 years old), occurs infrequently. Dentists, in particular, are not likely to address the matter in their practice.

Uncertainty about how to incorporate prevention and disease screening within clinical activities[11,12] and how to communicate disease prevention messages effectively[13] contribute to the deficiencies noted above. Schools of medicine and dental medicine have special obligations to educate students, residents, and postgraduates about prevention counseling. Training must emphasize the successively younger ages at which adolescents are experimenting with tobacco. Instruction in prevention skills is needed early and often in physicians' and dentists' curriculums. All young persons are at risk for tobacco addiction and are suitable targets for prevention counseling. □

Acknowledgments

The author acknowledges the assistance of the Connecticut Cancer Institute in completing this project, as well as the help of the Connecticut Chapter of the American Academy of Pediatrics, the Connecticut Academy of Family Physicians, and the Connecticut State Dental Association, which provided information on their memberships.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50212.
2. Diel HS. The physician and cigarette smoking. *New Physician.* 1970;19:231–234.
3. Seffrome JR, Stauffer DJ. Patient education on cigarette smoking: the dentist's role. *J Am Dent Assoc.* 1976;92:7514.
4. Epps RP, Manley MW. A physician's guide to preventing tobacco use during childhood and adolescence. *Pediatrics.* 1991;88:140–144.
5. Committee on Psychosocial Aspects of Child and Family Health. *Guidelines for Health Supervision II.* Elk Grove Village, Ill: American Academy of Pediatrics; 1988.
6. US Preventive Services Task Force. *Guide to Clinical Preventive Services.* Baltimore, Md: Williams & Wilkins; 1989:193.
7. Frank E, Winkleby MA, Altman DG, et al. Predictors of physicians' smoking cessation advice. *JAMA.* 1991;266:3139–3144.
8. Selected tobacco-use behaviors and dietary patterns among high school students—United States, 1991. *MMWR.* 1992;41:417–420.
9. Perry CL. Results of prevention programs with adolescents. *Drug Alcohol Depend.* 1987;20:13–19.
10. Kottke TE, Battista RN, DeFriese GH, et al. Attributes of successful smoking cessation interventions in medical practice: a meta analysis of 39 controlled trials. *JAMA.* 1988;259:2882–2889.
11. Carter WD, Belcher DW, Inui TS. Implementing preventive care in clinical practice, II: problems for managers, clinicians and patients. *Med Care Rev.* 1981;38:195–216.
12. Orlandi MA. Promoting health and preventing disease in health care settings: an analysis of barriers. *Prev Med.* 1987;16:119–130.
13. Orleans CT, George LK, Houpt JL, Brodie KH. Health promotion in primary care: a survey of U.S. family practitioners. *Prev Med.* 1985;14:636–647.

Preventing Firearm Injury among Children: Free Materials for Healthcare Professionals

On an average day, 14 children and teenagers die in unintentional and intentional gun incidents. The American Academy of Pediatrics, Dr C. Everett Koop, and the Center to Prevent Handgun Violence have created Steps to Prevent Firearm Injury (STOP), a program that provides materials for pediatric health professionals to use in counseling families on firearm injury prevention. The materials include information highlighting the danger of keeping a gun in the home and guidelines for safe home storage of guns and ammunition.

Pediatric health professionals may request a free kit containing 25 parent/patient brochures, 2 office posters, a monograph on background information, an audiotape with sample patient counseling sessions on gun safety, a counseling tip sheet, and suggestions for further reading.

To receive a kit, send your request to the Center to Prevent Handgun Violence, PO Box 425, Bladensburg, MD 20710-9974. For further information, contact Ricky O'Kane at the Center to Prevent Handgun Violence, tel (202) 289-7319.

An Increase in *Haemophilus influenzae* Type B Vaccination among Preschool-Aged Children in Inner-City Los Angeles, 1990 through 1992

Donnell P. Ewert, MPH, Suzanne Westman, MD, MPH, Bridget Ward, MS, RN, Nancie Bendaña, MS, RN, Ron Halbert, MD, MPH, Constance M. Vadheim, PhD, Joel I. Ward, MD, and Laurene Mascola, MD, MPH

ABSTRACT

Haemophilus influenzae type b (Hib) vaccination coverage and disease incidence were measured among preschool-aged children residing in inner-city Los Angeles. Among children 1.5 to 14 months of age, vaccination coverage of at least one dose increased from 0% in 1990 to 82% (95% confidence interval [CI] = 73%, 91%) in 1992. Among children 15 to 59 months old, vaccination coverage of at least one Hib dose administered at or after age 15 months increased from 35% (95% CI = 29%, 41%) in 1990 to 63% (95% CI = 56%, 70%) in 1992. Although Hib vaccination has reduced disease incidence in this population, greater use of vaccine can result in further reductions. (*Am J Public Health.* 1994;84:1154–1157)

Introduction

Before effective vaccines were widely administered, *Haemophilus influenzae* type b (Hib) was the leading cause of bacterial meningitis and an important cause of bacteremic infection in children less than 5 years of age. Invasive disease affected 1 in 200 children before the age of 5 years, causing death in 3% to 6% of cases and permanent sequelae in 20% to 30% of survivors.[1] Although the incidence was greatest in children less than 2 years old, effective Hib conjugate vaccines were not available for children as young as 18 months old until December 1987 and were not available for infants as young as 2 months of age until late 1990 (Table 1).[1,2]

The Immunization Demonstration Project, sponsored by the Centers for Disease Control and Prevention, was initiated during 1990 to improve vaccination coverage in a predominantly Latino-populated 53-census-tract area in inner-city Los Angeles. The project conducted annual community vaccination surveys from 1990 through 1992, which afforded an opportunity to examine changes in Hib vaccination coverage and disease incidence. Although precipitous declines of Hib disease have been documented within various populations,[2–7] no study, to our knowledge, has demonstrated both an increase in Hib vaccination coverage and a decline of Hib disease incidence within a geographic area.

Methods

The Immunization Demonstration Project established a walk-in immunization clinic at public comprehensive health center A, where no such clinic had previously existed. This new clinic administered 1139 Hib vaccine doses in 1990, 5660 doses in 1991, and 6294 doses in 1992. The project also increased the capacity of an already-established walk-in immunization clinic at public comprehensive health center B by increasing the nursing and clerical staff. Both health centers were located within the project catchment area.

The cluster survey methodology used extensively by the World Health Organization's Expanded Programme on Immunization was used to measure levels of Hib vaccination in preschool-aged children. This method has been used successfully in developing countries[8,9] and was recently adapted for use in urban areas of the United States.[10] During the three Immunization Demonstration Project surveys, 50 clusters were systematically selected from the 53-census-tract catchment area with probability proportionate to size. A random starting point was determined for each cluster by randomly selecting a city block and a residence on the block. Households were visited consecutively from the starting point, and all children from seven eligible households were included.

Household inclusion criteria were permanent residence of a child 6 weeks to 59 months of age and the presence of a respondent at least 16 years of age who was willing to participate. Possession of a written vaccination record was not required for inclusion in the survey but was required for inclusion in the analysis of

At the time of the study, Donnell P. Ewert was and Bridget Ward, Nancie Bendaña, and Laurene Mascola are with Acute Communicable Disease Control, Los Angeles County Department of Health Services. Suzanne Westman and Ron Halbert are with the Immunization Demonstration Project, John Wesley Community Health Institute, Los Angeles. Constance M. Vadheim and Joel I. Ward are with the Hib Surveillance Project, UCLA Center for Vaccine Research, Los Angeles.

Requests for reprints should be sent to Donnell P. Ewert, MPH, Communicable Disease Program, Indiana State Department of Health, 1330 W Michigan St, PO Box 1964, Indianapolis, IN 46206-1964.

This paper was accepted February 23, 1994.

Editor's Note. See related editorial by Shelov (p 1066) in this issue.

Hib vaccination. A standard interview form was used to obtain family information and the vaccination history of each child. While the type of vaccine provider (public or private) was recorded during all three surveys, the specific clinic that administered each vaccine dose was recorded only during the 1992 survey. The type of Hib vaccine received by each child was not determined (Hib capsular polysaccharide vaccine [polyribosyl-ribitol-phosphate, PRP], Hib diphtheria toxoid conjugate vaccine [PRP-D], Hib diphtheria CRM197 protein conjugate vaccine [HbOC], or Hib meningoccocal protein conjugate vaccine [PRP-OMP]). Bilingual (English/Spanish) interviewers conducted interviews between 4 PM and 7 PM on weekdays and between 9 AM and 5 PM on weekends. The surveys took place between April 18 and June 26 of each year.

Active surveillance for cases of Hib disease was performed throughout Los Angeles County from January 1986 through December 1992, except for the period between July 1987 and June 1988 when surveillance was passive. Active surveillance consisted of calling all hospital infection control practitioners and laboratories in Los Angeles County once or twice a month to request case reports, auditing a 50% sample of hospital laboratories to identify positive Hib cultures, and reviewing medical records of pediatric patients discharged with any *International Classification of Diseases*, Ninth Revision,[11] diagnosis compatible with Hib disease. Passive surveillance consisted of telephone and written reports to the Los Angeles County Department of Health Services from hospital infection control practitioners, laboratories, and physicians.

An Hib case was defined as involving a child less than 5 years of age who had (1) clinical findings consistent with invasive Hib infection plus positive cultures from a normally sterile site; (2) clinical signs and symptoms consistent with epiglottitis with positive cultures from the pharynx, epiglottis, or blood; or (3) clinical findings consistent with invasive Hib infection plus Hib antigen in fluid from a normally sterile body site. Population data from the 1980 census, updated for births and deaths that occurred during the intervening years, were used to calculate rates.

TABLE 1—Chronology of Hib Vaccine Licensure in the United States

Licensure Date	Vaccine Type	Age at Administration, mo
April 1985	Polysaccharide	24[a]
December 1987	Conjugate	18
April 1990	Conjugate	15
October 1990	Conjugate	2

[a]18 months for children enrolled in day care.

TABLE 2—Percentage of Children Vaccinated against Hib Disease in the Immunization Demonstration Project Catchment Area of Los Angeles, 1990 through 1992, by Dose and Age Group

Dose	Age, mo	1990			1991			1992		
		No.	%	95% CI	No.	%	95% CI	No.	%	95% CI
First[a]	1.5–14	94	0	. . .	111	32	21, 42	114	82	73, 91
Second[a]	4–14	73	0	. . .	88	0	. . .	95	70	60, 79
Third[a]	6–14	55	0	. . .	68	0	. . .	74	34	24, 44
15 mo[b]	15–59	303	35	29, 41	284	57	52, 63	283	63	56, 70

Note. CI = confidence interval.
[a]Infant dose.
[b]At least one dose at or after 15 months of age.

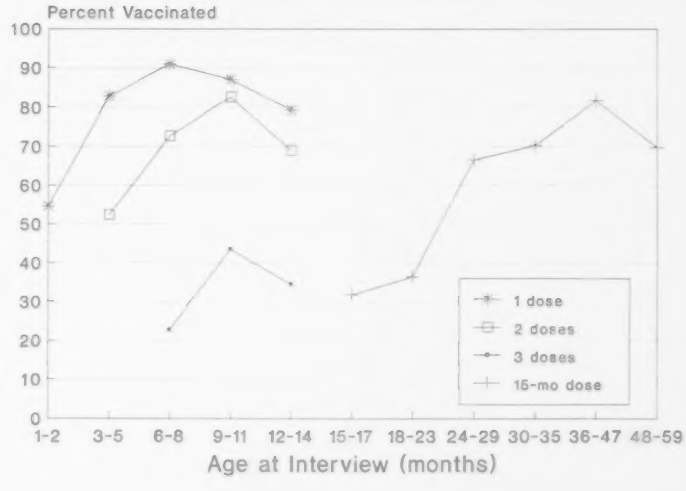


FIGURE 1—Percentage of children 6 weeks through 59 months of age residing in the Immunization Demonstration Project catchment area who had received Hib vaccine, by dose and age group: 1992.

Results

Only children with available written vaccination records were included in vaccination coverage estimates. Written records were available for 397 (79%) of 502 children in 1990, 395 (89%) of 445 children in 1991, and 397 (87%) of 458 children in 1992. The samples were 94% to 97% Latino and 2% to 6% African American; less than 1% of the children were members of other ethnic groups.

The proportion of children 1.5 to 14 months old who received at least one dose

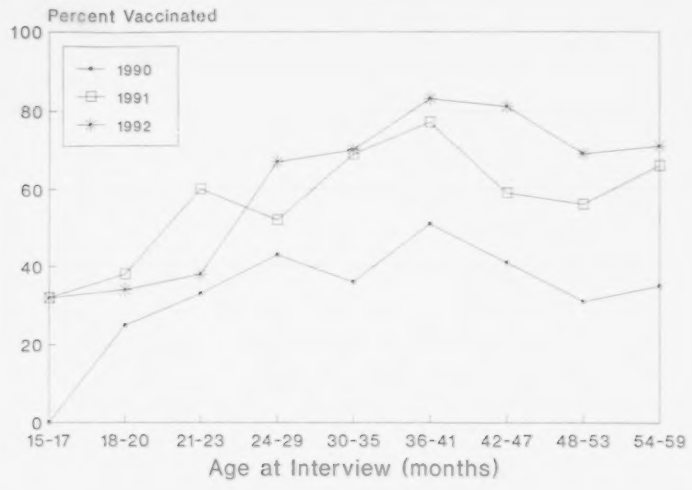


FIGURE 2—Percentage of children 15 to 59 months of age residing in the Immunization Demonstration Project catchment area who had received one dose of Hib vaccine after age 15 months, by age group: 1990 through 1992.

TABLE 3—Percentage of Children 15 to 59 Months of Age Who Received at Least One Dose of Hib Vaccine after Age 15 Months in the Immunization Demonstration Project Catchment Area of Los Angeles, 1990 through 1992, by Vaccine Provider

	1990			1991			1992		
Provider	No.	%	95% Confidence Interval	No.	%	95% Confidence Interval	No.	%	95% Confidence Interval
Health department	127	37	30, 45	133	59	50, 69	143	69	61, 78
Private sector	169	33	26, 41	139	56	48, 65	139	57	47, 67

TABLE 4—Incidence of Hib Disease in the Immunization Demonstration Project Catchment Area of Los Angeles, 1986 through 1989 Mean and 1990 through 1992, by Age Group

	Incidence (per 100 000 Person-Years)				
Age Group, mo	1986–1989 Mean	1990	1991	1992	Change,[a] %
<12	141.2	131.0	129.1	56.5	−60
12–23	84.5	78.0	15.3	30.1	−64
24–59	14.3	5.3	0	0	−100

[a]From 1986–1989 mean incidence to 1992 incidence.

of Hib vaccine before age 15 months increased significantly from 1990 to 1991 and from 1991 to 1992 (Table 2). By 1992, large proportions of children had been vaccinated with second and third doses of Hib vaccine before 15 months of age; coverage increased with age until 12 months of age but declined thereafter (Figure 1). The proportion of children 15 to 59 months old who received one Hib dose at or after 15 months of age increased significantly from 1990 to 1991; this proportion also increased from 1991 to 1992 but to a lesser and nonsignificant extent (Table 2). Hib vaccination coverage was most complete among children 36 to 41 months of age during all 3 years (Figure 2).

Vaccination coverage of the one Hib dose given at or after 15 months of age increased significantly from 1990 to 1991 among clients of both public and private providers (Table 3). Although not statistically significant, vaccination coverage continued to increase from 1991 to 1992 among children obtaining vaccine from the health department, while coverage did not increase among children using private sources. Of the 332 children enrolled in the 1992 survey who had received at least one dose of Hib vaccine at any age, 59 (18%) and 43 (13%) received at least one dose from public comprehensive health centers A and B, respectively.

In the Immunization Demonstration Project catchment area, Hib disease in children less than 5 years of age declined 69% from the 1986 through 1989 mean incidence of 55.9 (per 100 000 person-years) to the 1992 incidence of 17.6. Incidence remained greatest among infants (Table 4) but declined 56% from the 1991 incidence of 129.1 to the 1992 incidence of 56.5. No children 24 through 59 months of age with Hib disease were reported in the project catchment area from February 1990 through December 1992.

Discussion

Dramatic improvement in Hib vaccination coverage among children less than 15 months of age occurred in this catchment area from 1990 to 1992. It is notable that Hib disease in preschool-aged children declined almost 70% from the 1986 through 1989 mean incidence to the 1992 incidence, even without active surveillance from July 1987 through June 1988, which was part of the baseline time period. The incidence of Hib disease in infants began to decrease before the introduction of infant vaccines in 1990, a result of vaccinating older children and reducing oropharyngeal carriage.[12] However, the precipitous decline in incidence among infants occurred during 1992, after infant vaccine was licensed. Although

vaccination against Hib disease among infants has increased rapidly, Hib vaccination coverage during the first year of life was still less complete during 1992 than coverage of infant doses of diphtheria–tetanus–pertussis (DTP) vaccine and oral polio vaccine (OPV) (Immunization Demonstration Project, unpublished data, 1992). Since incidence can be reduced further in infants by more timely and complete vaccination, improvement in Hib vaccination coverage during the first year of life should continue to be the main focus of public health interventions against Hib disease.

Vaccination coverage in the Immunization Demonstration Project catchment area increased significantly from 1990 to 1991 among children 15 to 59 months old. This increase, which probably resulted from more widespread use of Hib vaccine by providers, raised coverage to the level commonly seen with the fourth dose of DTP and the third dose of OPV. The small improvement in coverage from 1991 to 1992, especially among children vaccinated in the private sector, suggests that further improvement in coverage of the one Hib dose received at or after 15 months of age may be as difficult as improving coverage of other vaccine doses administered during the second year of life.

There were several limitations to this study. First, since the type of Hib vaccine received was not recorded during data collection and only three doses of PRP-OMP are required (with the third dose given at age 12 months), a small number of PRP-OMP final doses may have been misclassified as third doses in the 1992 survey. Second, since the data were analyzed by age group for different doses and children were not evenly distributed by age across all clusters, some clusters had more children represented than other clusters for the estimates of infant coverage, while the representation was opposite for the coverage estimate among older children. This unequal weighting may have introduced some unknown bias into the coverage estimates. Third, modifying the Expanded Programme on Immunization methodology to include seven households, instead of seven children, may have resulted in oversampling of households with more than one preschooler. Fourth, by attempting to limit recall bias by excluding from the analysis the 10% to 20% of children without written vaccination records, vaccination coverage may have been overestimated since respondents for children without records reported lower vaccination levels. This overestimation was probably very small because only small portions of the samples were excluded, and parents of children without written records tend to underestimate the number of doses previously received.[13] Finally, if persons not at home during the survey hours were different from those at home, some unknown selection bias may have been introduced into the coverage estimates. □

Acknowledgments

This study was supported in part by Centers for Disease Control and Prevention grant 666/CCU 901919-02, National Institutes of Health grant R01-AI-31580, and grants from Connaught Laboratories Inc, Merieux Institute Inc, and Merck Sharp & Dohme.

We wish to thank Leslie Hemenway, Eileen Eriksen, and Lawrence Valle of the Hib Surveillance Project for their assistance with active surveillance and Hector Bedoya, Patricia Luppi, and Arturo Picazo from the Immunization Demonstration Project for their assistance with the cluster surveys.

References

1. Centers for Disease Control. *Haemophilus* b conjugate vaccines for prevention of *Haemophilus influenzae* type b disease among infants and children two months of age and older: recommendations of the Immunization Practices Advisory Committee (ACIP). *MMWR.* 1991;40(RR-1):1–7.
2. Adams WG, Deaver KA, Cochi SL, et al. Decline of childhood *Haemophilus influenzae* type b (Hib) disease in the Hib vaccine era. *JAMA.* 1993;269:221–226.
3. Santosham M, Wolff M, Reid R, et al. The efficacy in Navajo infants of a conjugate vaccine consisting of *Haemophilus influenzae* type b polysaccharide and *Neisseria meningitidis* outer-membrane protein complex. *N Engl J Med.* 1991;324:1767–1772.
4. Black SB, Shinefield HR, Kaiser Permanente Pediatric Vaccine Study Group. Immunization with oligosaccharide conjugate *Haemophilus influenzae* type b (HbOC) vaccine on a large health maintenance organization population: extended follow-up and impact on *Haemophilus influenzae* disease epidemiology. *Pediatr Infect Dis J.* 1992;11:610–613.
5. Vadheim CM, Greenberg DP, Eriksen E, et al. Eradication of Hib disease in Southern California. *Arch Pediatr Adolesc Med.* 1994;148:51–56.
6. Broadhurst LE, Erickson RL, Kelley PW. Decreases in invasive *Haemophilus influenzae* diseases in US Army children, 1984 through 1991. *JAMA.* 1993;269:227–231.
7. Murphy TV, White KE, Pastor P, et al. Declining incidence of *Haemophilus influenzae* type b disease since introduction of vaccination. *JAMA.* 1993;269:246–248.
8. Henderson RH, Sundaresan T. Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. *Bull WHO.* 1982;60:253–260.
9. Lemeshow S, Stroh G Jr. *Sampling Techniques for Evaluating Health Parameters in Developing Countries.* Washington, DC: National Academy Press; 1988:8–13.
10. Ewert DP, Thomas JC, Chun LY, Enguidanos RC, Waterman SH. Measles vaccination coverage among Latino children aged 12 to 59 months in Los Angeles County: a household survey. *Am J Public Health.* 1991;81:1057–1059.
11. National Center for Health Statistics. *International Classification of Diseases, 9th Revision, Clinical Modification.* Ann Arbor, Mich: Commission on Professional and Hospital Activities; 1990.
12. Takala AK, Eskola J, Leinonen M, et al. Reduction of oropharyngeal carriage of *Haemophilus influenzae* type b (Hib) in children immunized with an Hib conjugate vaccine. *J Infect Dis.* 1991;164:982–986.
13. Valadez JJ, Weld LH. Maternal recall error of child vaccination status in a developing nation. *Am J Public Health.* 1992;82:120–122.

Using a Geographic Information System to Understand Child Pedestrian Injury

Mary Braddock, MD, MPH, Garry Lapidus, PA-C, MPH, Ellen Cromley, PhD, Robert Cromley, PhD, Georgine Burke, PhD, and Leonard Banco, MD

ABSTRACT

Data from police accident reports involving pedestrians less than 20 years of age in Hartford, Conn, during 1988 through 1990 were abstracted and entered into a geographic information system. Two high-frequency collision areas were identified and compared. There were 374 child pedestrians involved in collisions (a rate of 28 per 10 000). Two high-occurrence areas accounted for 30% of collisions. Collisions in one of these areas were more likely to involve younger children (8.1 vs 10.2 years of age) and to occur in the late afternoon, and occurred closer to the child's residence, than collisions in the other area. The geographic information system is a useful tool in the study of child pedestrian collisions. (*Am J Public Health.* 1994;84:1158–1161)

Introduction

Mapping disease events is a traditional approach to public health research.[1–7] However, point maps are an inefficient way of providing information for the development, implementation, and evaluation of public health programs.[8]

Geographic information systems allow management, analysis, and reporting of thematic data (demographics, health status indicators) within their geographic context (locations, boundaries).[9] Because many injuries have a strong geographic component, they are well suited to geographic information system analysis.[10-18] The purpose of this study was to demonstrate the usefulness of a geographic information system in examining child pedestrian collisions in Hartford, Conn.

Methods

All motor vehicle collisions reported to the Hartford police department between January 1, 1988, and December 31, 1990, involving pedestrians under 20 years of age were examined. Copies of police accident reports were obtained from the Connecticut Department of Transportation, and the following information was abstracted: date and time of collision, collision location, environmental conditions, vehicle type, pedestrian/vehicle maneuvers, contributing factors, pedestrian/driver age and sex, pedestrian place of residence, and injury severity.

A commercial geographic information system, TransCad,[19] was used to code the site of collision and pedestrian residence by address-matching these sites with the 1990 US Census TIGER/Line files for Hartford County.[20] Distances from collision sites to pedestrian residences were computed. The geographic information system produced point maps of collision and pedestrian residential locations.

Two areas of high collision density were identified for analysis. Descriptive, univariate statistics for collision, driver, and pedestrian characteristics were calculated. Because the data include the universe of all collisions during 1988 through 1990, inferential statistical procedures were not performed. Population data used to calculate rates were obtained from the 1990 census.[21]

Results

Between January 1, 1988, and December 31, 1990, there were 358 child pedestrian collisions in Hartford. These collisions occurred uniformly throughout the study period and involved 382 vehicles and 374 child pedestrians, resulting in an annual age-specific collision rate of 28 per 10 000 persons.

A point map of child pedestrian collision locations reveals that the collisions were distributed in a nonuniform manner with several high-frequency areas (Figure 1). Figure 2 selectively illustrates two high-frequency collision areas that accounted for 30% of all collisions. The Albany Avenue area is a 1.4 × 0.8 km (0.9 × 0.5 mile) section along Albany Avenue, a major commuting artery carrying an average traffic volume of 10 000 to 25 000 motor vehicles per day. The Park Street area is a 1.3 × 0.5 km (0.8 × 0.3 mile) section of a collector road carrying an average traffic volume of 3000 to 10 000 vehicles per day.[22] Both neighborhoods are densely populated with two- to three-story buildings of mixed residential/commercial land use.

Mary Braddock, Garry Lapidus, Georgine Burke, and Leonard Banco are with the Connecticut Childhood Injury Prevention Center, Hartford. Mary Braddock is also with St. Paul's Children's Hospital, St. Paul, Minn. Garry Lapidus, Georgine Burke, and Leonard Banco are also with the Department of Pediatrics, Hartford Hospital, and the University of Connecticut, Farmington. Ellen Cromley and Robert Cromley are with the Department of Geography, University of Connecticut, Storrs.

Correspondence should be sent to Garry Lapidus, PA-C, MPH, Connecticut Childhood Injury Prevention Center, 80 Seymour St, Hartford, CT 06102-5037.

This paper was accepted February 15, 1994.

FIGURE 1—Child pedestrian–motor vehicle collision sites: Hartford, Conn, 1988 through 1990.

FIGURE 2—Albany Avenue (+) and Park Street (open circles) area collision sites, 1988 through 1990.

Sixty-nine collisions occurred in the Albany Avenue area, and 38 collisions occurred within the Park Street area. Figure 3 illustrates the location of the residences of children involved in pedestrian collisions in the two areas.

The characteristics of the two collision areas are compared in Table 1. In comparison with collisions in the Albany Avenue area, a greater proportion of those in the Park Street area occurred in the hours after school dismissal (61% vs 46%), on weekends (29% vs 20%), and during the summer (32% vs 17%). Children struck in the Albany Avenue area were older (mean ± SD = 10.2 ± 5 vs 8.1 ± 5 years) and more likely to be seriously or fatally injured (32% vs 13%) than those struck in the Park Street area. Driver error was more often a factor in the collisions occurring in the Albany Avenue area (29% vs 18%). Collisions in the Albany Avenue area also occurred at a greater (mean ± SD) distance from the child's place of residence (0.62 ± 0.8 vs 0.43 ± 0.8 km [0.39 ± 0.5 vs 0.27 ± 0.5 miles]). Collisions in the Park Street area more often occurred at intersections without a traffic signal (30% vs 14%). Collisions in both areas tended to occur on dry pavement and to involve male drivers in passenger cars. In general, characteristics of collisions in the Albany Avenue area were more similar to those elsewhere in the city.

Discussion

Pedestrian collisions are a leading cause of injury death and hospitalization among school-aged children.[12-14,23] Research shows that these collisions are not uniformly distributed within a city.[12-14] This study used a geographic information system to illustrate differences between two collision sites in Hartford, Conn.

Data were taken from police accident reports, which are known to underestimate the occurrence of nonroadway pedestrian collisions.[24,25] Nevertheless, police data include most serious traffic injuries to children and contain other valuable information concerning the collision event.[12-14,24-26]

This study found several differences between the two collision sites in Hartford. Of particular interest was the differing proximity of child residences to pedestrian collision sites. This type of information, which would not be easily accessible without geographic information system technology, has important implications for understanding and preventing child pedestrian collisions. A site-specific intervention suitable for the Albany Avenue area might include education aimed at school-aged children; environmental modifications designed to decrease traffic flow, improve intersection crossing times, and increase child pedestrian visibility; and improved police enforcement of vehicle speed laws and adherence to traffic signs. In the Park Street area, educational efforts directed at parents of preschoolers and school-

FIGURE 3—Residential locations of children involved in pedestrian collisions in the Albany Avenue (+) and Park Street (open circles) areas, 1988 through 1990.

TABLE 1—Attributes of Collisions by Site: Hartford, Conn, 1988 through 1990

	Albany Ave (n = 69)	Park St (n = 38)	All Other (n = 251)
Environment			
Season, %			
Summer	17	32	30
Fall	33	26	23
Winter	13	16	16
Spring	36	27	32
Time of day, %			
6 AM–10 AM	12	0	9
10 AM–2 PM	20	18	15
2 PM–6 PM	46	61	48
6 PM–10 PM	20	16	27
Weekend day (Saturday, Sunday), %	20	29	30
Slippery road conditions (rain, snow, ice, oil), %	17	13	17
Driver			
Mean age, y (SD)	32.7 (11)	35 (14)	35.5 (15)
Male, %	74	79	73
Passenger car, %	83	87	83
Driver contributing factors (inattentiveness, speed, traffic violation), %	29	18	29
Nonturning vehicle manuever, %	71	84	78
Pedestrian			
Mean age, y (SD)	10.2 (5)	8.1 (5)	10.0 (5)
Median age, y	9	7	9
Age distribution, y, %			
0–4	9	29	13
5–9	49	40	40
10–14	17	18	21
15–19	25	13	26
Male, %	58	58	67
Pedestrian contributing factors, %			
Crossing between intersections	47	40	46
Crossing at intersection			
With no traffic signal	14	30	13
Against traffic signal	5	4	5
Walking/playing in roadway	3	5	5
Crossing behind other vehicle	19	16	19
Distance from home to collision site, km, mean (SD)	0.62 (0.8)	0.39 (0.8)	0.82 (1.1)
Injury severity, %			
Deat	3	0	1
Disa ity	29	13	29
Visible, not disabling	33	53	32
Probable, not visible	33	32	35
No injury	2	2	3

aged children, increased use of warning signs and crossing signals, and the development of off-street playground areas might be effective prevention strategies.

We are currently investigating other geographic features that may play an important role in child pedestrian collision etiology, such as schools, parks, playgrounds, and convenience stores. In conclusion, the ability of a geographic information system to illustrate spatial relationships between pedestrian collision sites and other city landmarks advances the study of pedestrian injuries. □

Acknowledgments

We thank David Gregorio, PhD, MS, and Robert Zavoski, MD, MPH, for their help in reviewing the manuscript. We also thank Gene Interlandi, Harley Polverelli, and Mario Tonarelli from the Connecticut Department of Transportation for their assistance in providing data and the University of Connecticut Research Foundation for its support.

References

1. Nimmo AW. The chief scientist reports . . . review of computer mapping of health data. *Health Bull.* 1989;47(1):40–48.
2. Frazier TM, Lalich NR, Pederson DH. Uses of computer-generated maps in occupational hazard and mortality surveillance. *MMWR.* 1984;33(suppl):27S–32S.
3. Breckenridge MB, Tallia AF, Like RC. Display of small-area variation in health-related data: a methodology using resistant statistics. *Soc Sci Med.* 1988;26:141–151.
4. Schwartz T, Callen JP, Silva J Jr. A cluster

of Hodgkin's disease in a small community—evidence for environmental factors. *Am J Epidemiol.* 1978;108:19–26.
5. Abel U, Becker N. Geographical clusters and common patterns in cancer mortality in the Federal Republic of Germany. *Arch Environ Health.* 1987;42:51–57.
6. Rip M, Bourne D. The spatial distribution of infant mortality rates in South Africa, 1982. *S Afr Med J.* 1988;73:224–226.
7. Lam N. Geographical patterns of cancer mortality in China. *Soc Sci Med.* 1986;23: 241–247.
8. Rothman KJ. Clustering of disease. *Am J Public Health.* 1987;77:13–15.
9. Goodchild MF. A spatial analytic perspective on geographic information systems. *Int J Geogr Information Syst.* 1987;1:327–334.
10. *Injury Mortality Atlas of the United States, 1979–1987.* Atlanta, Ga: Centers for Disease Control; 1991.
11. Waller AE, Baker SP, Szocka A. Childhood injury deaths: national analysis and geographic variations. *Am J Public Health.* 1989;79:310–315.
12. Braddock M, Lapidus G, Gregorio D, et al. Population, income, and ecological correlates of child pedestrian injury. *Pediatrics.* 1991;88:1242–1247.
13. Rivara F, Barber M. Demographic analysis of childhood pedestrian injury. *Pediatrics.* 1985;76:375–381.
14. Lapidus G, Braddock M, Banco L, et al. Child pedestrian injury: a population collision and injury severity profile. *J Trauma.* 1991;31:1110–1115.
15. Speigel CN, Lindaman FC. Children can't fly: a program to prevent childhood morbidity and mortality from window falls. *Am J Public Health.* 1977;67:1143–1147.
16. Centers for Disease Control. Queens Boulevard Pedestrian Safety Project—New York City. *MMWR.* 1989;38:61–63.
17. Fingerhut LA, Ingram DD, Feldman JJ. Firearm homicide among black teenage males in metropolitan counties. Comparison of death rates in two periods, 1983 through 1985 and 1987 through 1989. *JAMA.* 1992;267:3054–3058.
18. Fingerhut LA, Ingram DD, Feldman JJ. Firearm and nonfirearm homicide among persons 15 through 19 years of age. Differences by level of urbanization, United States, 1979 through 1989. *JAMA.* 1992;267: 3048–3053.
19. *TransCad, Version 2.0.* Newton, Mass: Caliper Corp; 1990.
20. *TIGER/Line Census Files, 1990* [machine-readable data files]. Washington, DC: US Bureau of the Census; 1991.
21. State of Connecticut. *1990 Census of Population.* Hartford, Conn: Connecticut Census Data Center, Office of Policy and Management; 1991.
22. *Comprehensive Plan of Development, Land Use Component.* Hartford, Conn: City of Hartford Commission on the City Plan; 1985.
23. Lapidus G, Banco L. Childhood injuries in Hartford, Connecticut. *Conn Med.* 1990; 54(2):51–55.
24. Agran P, Castillo D, Winn D. Limitations of data compiled from police reports on pediatric pedestrian and bicycle motor vehicle events. *Accident Anal Prev.* 1990;22: 361–370.
25. Meuller B, Rivara F, Bergman A. Factors associated with pedestrian-vehicle injuries and fatalities. *West J Med.* 1987;146:243–245.
26. Pless IB, Verreault R, Arsenault L, et al. The epidemiology of road accidents in childhood. *Am J Public Health.* 1987;77:358–360.

The Compliance of Licensed US Child Care Centers with National Health and Safety Performance Standards

David G. Addiss, MD, MPH, Jeffrey J. Sacks, MD, MPH, Marcie-jo Kresnow, MS, Joann O'Neil, BS, and George W. Ryan, PhD

ABSTRACT

The American Public Health Association and the American Academy of Pediatrics recently published health and safety guidelines for child care centers. A survey was conducted to determine the extent to which practices in US child care centers are reflective of these guidelines. Compliance with 16 guidelines ranged from 19.5% to 98.6%, varied considerably by state, and was not consistently associated with selected center characteristics. Prevention efforts should focus on practices for which compliance is low and on those that have the greatest disease- and injury-reducing potential. (*Am J Public Health.* 1994;84:1161–1164)

Introduction

In 1992, the American Public Health Association and the American Academy of Pediatrics published performance guidelines for the control of infectious diseases and injuries in US child care centers.[1] Adoption of these guidelines by providers will depend, in part, on whether compliance requires substantial or costly changes in child care practices, policies, or facilities. In November and December 1990, before the guidelines were published, we conducted a survey to determine the extent to which policies and practices in US child care centers were consistent with these guidelines.

Methods

A total of 2003 child care center directors were interviewed by telephone about their policies and practices for control and prevention of infectious diseases and injuries. This nationally representative sample was selected from lists of licensed child care centers in all 50 states, 14 local jurisdictions, and the District of Columbia. A minimum of 20 centers was randomly selected from each jurisdiction; the remainder of the sample was proportionately allocated according to the number of centers in each jurisdiction. Responses were weighted to obtain national estimates, and SUDAAN software[2] was used to compute 95% confidence inter-

David G. Addiss is with the National Center for Infectious Diseases, and Jeffrey J. Sacks, Marcie-jo Kresnow, Joann O'Neil, and George W. Ryan are with the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, Ga.

Requests for reprints should be sent to David G. Addiss, MD, MPH, Division of Parasitic Diseases (F-22), Centers for Disease Control and Prevention, 4770 Buford Hwy, NE, Atlanta, GA 30341.

This paper was accepted February 4, 1994.

TABLE 1—Percentage of Child Care Centers in Compliance with Recommended Practices and Policies to Control Infectious Diseases and Injuries

Performance Guidelines Reference	Practice or Policy Recommended by Guidelines	Centers Reporting Recommended Practice	
		%	95% CI
	Infectious disease practices		
HP 31	Staff use single-use disposable towels to dry hands	98.6	98.0, 99.1
HP 31	4-year-old children use paper towels to dry hands	98.2	97.6, 98.8
HP 79	Center has written policies for management of ill children	93.8	92.7, 94.9
FA 149	All diaper changing surfaces adjacent to a sink	76.1	73.8, 78.5
HP 25	Staff wash hands after every diaper change	73.7	71.2, 76.2
HP 24	Diaper changing surfaces sanitized after every diaper change	73.0	70.5, 75.5
HP 80	Director required to report a case of measles to a health agency	65.6	63.5, 67.8
HP 29	4-year-old children always wash hands after toileting	40.0	37.7, 42.4
ID 23f	Director required to report an outbreak of diarrhea (≥ 3 cases in children attending the same classroom) to a health agency	19.5	17.6, 31.3
	Injury control practices		
AD 70	Document all injuries serious enough to require first aid	89.8	88.4, 91.2
AD 34	Practice fire drills at least once a month	75.5	73.7, 77.3
AD 70	Injury reports filed in a separate log	48.5	46.2, 50.9
FA 251	Maximum height of playground equipment ≤ 165 cm (5.5 ft)	45.2	42.8, 47.6
App 33(o)	Poison control number posted by telephone	42.0	39.6, 44.4
FA 245	Surface under highest piece of playground equipment adequately resilient based on surface and on depth and height of equipment[a]	37.8	35.4, 40.2
ST 44	All staff certified in first aid	36.0	33.9, 38.2

Note. Figures are based on reports by child care center directors in a telephone interview. A copy of the questionnaire used in the interview is available from the authors. Performance references refer to recommendations included in the American Public Health Association/American Academy of Pediatrics guidelines.[1]

[a]The following were considered not in compliance: asphalt or concrete playground surfaces, equipment ≥ 120 cm (4 ft) above bare earth, equipment ≥ 240 cm (8 ft) above synthetic mat surfaces, equipment ≥ 120 cm (4 ft) above a loose fill surface (e.g., bark, wood chips, gravel, sand) ≤ 2.54 cm (1 in) deep, equipment ≥ 150–180 cm (5–6 ft) over loose fill ≤ 15.24 cm (6 in) deep, and equipment ≥ 300–360 cm (10–12 ft) above specific surfaces (details available on request from the authors).

vals. Logistic regression models were run to assess center characteristics associated with reported compliance with each of the 14 practices for which compliance was less than 95%. Each model included six center characteristics: number of children (< 40, 40 through 74, ≥ 75), average weekly fee per child (< $50, $50 through $69, ≥ $70), region of the country (South, Northeast, Midwest, West),[3] for-profit status, independent ownership (i.e., not part of a local, regional, or national chain), and affiliation with a religious institution that influenced center policies (no, yes). The Wald test was used to assess whether compliance with recommended practices was associated with any single characteristic after adjusting for the five other characteristics in the model.

Results

The response rate was 85%.[4] Child care centers cared for a median of 40 children full-time; the median weekly fee was $59 per child (range = $0 to $345). Forty-seven percent of centers were not-for-profit organizations, 74% were owned independently, and 31% were affiliated with a religious institution. Diapered children were accepted by 69% of centers.

Compliance with recommended practices varied (Table 1). Seventy-four percent of the directors of child care centers with diapered children reported that staff washed their hands after every diaper change. In contrast, 40% of all directors reported that 4-year-old children in their center always washed their hands after toileting. Directors in 94% of centers indicated that they had written policies specifying how mildly ill children should be managed; 66% stated that they were required to report a case of measles to a health agency or official, but only 20% stated that they were required to report an outbreak of diarrhea.

Almost 90% of directors indicated that they documented all injuries serious enough to require first aid, and 49% of centers filed injury reports in a separate log (Table 1). Emergency evacuation drills were practiced at least monthly by 76% of centers. Directors were asked to describe the surface under the highest piece of playground equipment; 38% of surfaces were considered to be adequately resilient. The poison control telephone number was reportedly posted by the telephone in 42% of centers.

Independent ownership, religious affiliation, and for-profit status of the child care center were associated with decreased compliance (odds ratio < 1) with some recommended practices (Table 2). The direction of association between recommended practices and other center characteristics, including fee and number of children, was inconsistent. Although region of the country was a significant predictor for more than half of the practices considered, the specific region that predicted compliance varied. We observed considerable state-to-state variability in performance of recommended practices within each region, and, even within individual states, reported levels of compliance were high for some practices and low for others (Figure 1).

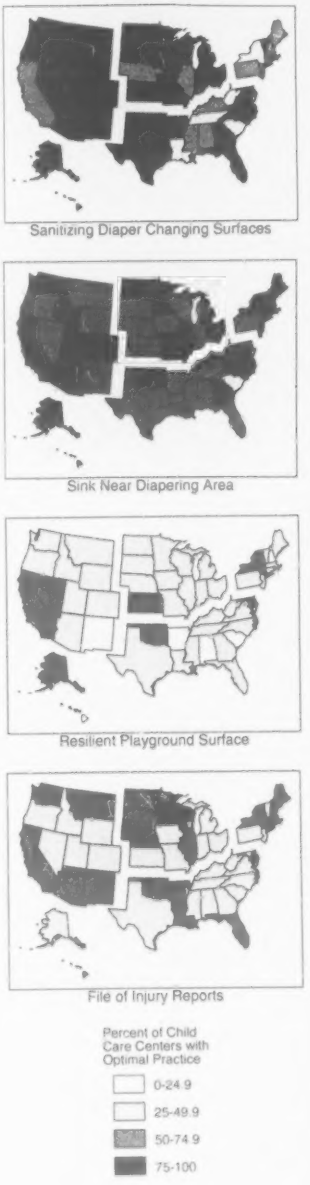
Discussion

Compliance with selected infectious disease and injury control practices recommended by the American Public Health Association and the American Academy of Pediatrics varies widely among licensed child care centers in the United States. Some recommended practices have been widely adopted, such as drying hands with single-use disposable towels, while others have generally not been implemented, such as reporting outbreaks of diarrhea to a state or local health agency. In general,

TABLE 2—Association between Child Care Center Characteristics and Compliance with Recommended Injury and Infectious Disease Policies and Practices

Recommended Practice	Region	Increasing Fee	Increasing No. of Children	Independent Ownership	Religiously Affiliated	For-Profit Status
Infectious diseases						
Written policy on ill children			↑ ↑			
Sink near diaper change areas		↑			↓ ↓	
Staff hand washing						
Sanitizing diaper changing surface	**	↑ ↑		↓ ↓		
Reporting measles	**	↑ ↑				
Child hand washing	*					↓
Reporting diarrhea outbreak	*			↓ ↓		
Injuries						
Documentation of injury	**		↑ ↑			↓ ↓
Fire drill practice	**					
Filing of injury reports					↓	
Playground equipment height	**		↓ ↓			
Posting poison control number	**					
Resilient playground surface		↑				
First aid training		↓	↓ ↓			

Note. Figures are based on reports by child care center directors in a telephone interview. The referents were as follows: southern region, fee < $50, < 40 children, nonindependent ownership, not affiliated with a religious institution, and not-for-profit status. The symbols ** and * indicate that compliance within one region significantly differed from that of other regions ($P < .01$ and $P < .05$, respectively). For dichotomous variables, ↑ ↑ and ↑ indicate an odds ratio > 1 compared with referent group ($P < .01$ and $P < .05$, respectively); ↓ ↓ and ↓ indicate an odds ratio < 1 compared with referent group ($P < .01$ and $P < .05$, respectively). For the continuous variables fee and number of children, ↑ ↑ and ↑ indicate a pattern of increasing odds ratios ($P < .01$ and $P < .05$, respectively), while ↓ ↓ and ↓ indicate a pattern of decreasing odds ratios ($P < .01$ and $P < .05$, respectively), as fee or number of children increases. For example, the proportion of directors reporting sinks near all diaper changing areas increased with increasing fee ($P < .05$) and was lower for religiously affiliated centers ($P < .01$).



Note. For each recommended practice, standard errors of state-specific point estimates exceeded 10% in as many as 15 states.

FIGURE 1—Percentage of child care centers complying with four selected infectious disease and injury control policies and practices, by state.

higher compliance was reported for infection control practices than for injury control practices, and a high degree of compliance was reported for procedures that could be easily observed or validated by licensing officials during inspections, such as having written policies for the care of ill children. However, performance was also low for some practices that could easily be verified, such as posting the poison control number by the telephone.

Implementation of some of the recommended policies and practices will require substantial changes in behavior or facilities. Performance of recommended practices was not consistently associated with specific child care center characteristics; thus, at a national level, these characteristics are unlikely to help identify high-risk centers in need of special interventions by licensing and public health officials. State-to-state variations among centers suggest that other factors such as differences in child care licensing, regulation, or enforcement[5,6]; staff training or availability of technical consultation[7]; or economic resources may be more strongly associated with performance of recommended practices.

Our study has several limitations. Actual practices could not be verified. Some interview questions may not have accurately measured compliance with a particular performance standard or may have addressed only one component of that standard. For example, the American Public Health Association/American Academy of Pediatrics guidelines recommend that handwashing sinks be "adjacent to diaper changing tables," whereas center directors were asked whether all diaper changing surfaces were "within an

arm's length (3 feet) of a sink for hand washing." The survey was conducted 1.5 years before the guidelines were published, but we believe it unlikely that practices changed significantly during this time. Finally, the precision of some state-specific estimates of reported practices was low; the standard error was greater than 10% for as many as 15 states.

Although considerable variation exists among child care centers, current infectious disease and injury control policies and practices in these facilities fall short of the American Public Health Association/American Academy of Pediatrics guidelines. Efforts should focus on improving policies and practices for which compliance is low and for which there is a demonstrated potential for reducing child care–related disease and injury. For example, only 38% of centers reported having adequately resilient surfaces on the playground; increasing this percentage could reduce the rate of injury.[8] Similarly, increased hand washing by children after toileting, currently estimated at only 40%, could contribute to reduced rates of diarrheal disease.[9] These national estimates provide baseline data for measuring the impact of efforts to reduce illness and injury in US child care centers. □

Acknowledgments

This research was presented, in part, at the 119th Annual Meeting of the American Public Health Association, November 1991, Atlanta, Ga.

We thank the participating day-care center directors, members of an expert panel that provided initial guidance, Dr Dennis Juranek for helpful review, and Susan Zaro, Thomas Chapel, Lela Baughman, Ed Hale, Nicole Lezin, George Latzanich, and Michael Errecart of Macro Systems, International.

References

1. *Caring for Our Children: National Health and Safety Performance Standards: Guidelines for Out-of-Home Child Care Programs.* Washington, DC/Elk Grove Village, Ill: American Public Health Association/American Academy of Pediatrics; 1992.
2. Shah BV. *Software for Survey Data Analysis (SUDAAN) Version 5.30.* Research Triangle Park, NC: Research Triangle Institute; 1989.
3. Trends in hospital utilization, United States, 1965–1985. *Vital Health Stat [13].* 1989; no. 101. DHHS publication PHS 89-1762.
4. Briss PA, Sacks JJ, Addiss DG, Kresnow M, O'Neil J. A nationwide study of the risk of injury associated with day care center attendance. *Pediatrics.* 1994;93:364–368.
5. Runyan CW, Gray DE, Kotch JB, Kreuter MW. Analysis of US child care safety regulations. *Am J Public Health.* 1991;81:981–985.
6. Morgan GG, Stevenson CS, Fiene R, Stephens KO. Gaps and excesses in the regulation of child day care: report of a panel. *Rev Infect Dis.* 1986;8:634–643.
7. Aronson SS, Aiken LS. Compliance of child care programs with health and safety standards: impact of program evaluation and advocate training. *Pediatrics.* 1980;65:318–325.
8. Sacks JJ, Holt KW, Holmgreen P, Colwell LS, Brown JM. Playground hazards in Atlanta child care centers. *Am J Public Health.* 1990;80:986–988.
9. Black RE, Dykes AC, Anderson KE, et al. Handwashing to prevent diarrhea in day-care centers. *Am J Epidemiol.* 1981;113:445–451.

Hearing Impairment among 10-Year-Old Children: Metropolitan Atlanta, 1985 through 1987

Carolyn D. Drews, PhD, Marshalyn Yeargin-Allsopp, MD, Catherine C. Murphy, MPH, and Pierre Decouflé, ScD

ABSTRACT

The prevalence of hearing impairment among 10-year-old children in metropolitan Atlanta between 1985 and 1987 was evaluated. Hearing-impaired children were identified by reviewing records at public schools and health and social service agencies. The prevalence was 1.1 per 1000 and was slightly higher among Blacks and boys than among Whites and girls. The most common known causes of hearing impairment were meningitis (0.3 per 1000), genetic and hereditary conditions (0.2 per 1000), and congenital rubella syndrome (0.1 per 1000). For 55% of the children, the etiology of the hearing loss could not be determined. Most (74%) of the children were diagnosed after the age of 2, suggesting that methods of early identification need to be improved. (*Am J Public Health.* 1994;84:1164–1166)

Introduction

Little information is available on the occurrence of childhood hearing impairment. In this paper we examine three aspects of childhood hearing impairment in metropolitan Atlanta during the years 1985 through 1987: its prevalence, its causes, and its association with other developmental disabilities.

Methods

The Metropolitan Atlanta Developmental Disabilities Study estimated the prevalence of five developmental disabilities among children who were born in 1975, 1976, or 1977 and who were living in metropolitan Atlanta when they were 10 years old.[1] This analysis focuses on hearing impairment, defined as an unaided bilateral pure tone hearing loss averaging at least 40 dB in the better ear at 500, 1000, and 2000 Hz.

Hearing-impaired children were identified by reviewing records of public schools, hospitals, and health and social service agencies, including all programs for the hearing-impaired sponsored by the state of Georgia. Data from all sources were reviewed to maximize ascertainment and to obtain complete information on

Carolyn D. Drews is with the Division of Epidemiology, Emory University School of Public Health, and the Department of Ophthalmology, Emory University School of Medicine, Atlanta, Ga. Marshalyn Yeargin-Allsopp, Catherine C. Murphy, and Pierre Decouflé are with the Division of Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Ga. Catherine C. Murphy is also with the Battelle Memorial Institute, Arlington, Va.

Requests for reprints should be sent to Carolyn D. Drews, PhD, Division of Epidemiology, Emory University School of Public Health, 1599 Clifton Rd, NE, Atlanta, GA 30329.

This paper was accepted November 4, 1993.

TABLE 1—Prevalence of Hearing Impairment (per 1000) among 10-Year-Old Children Living in Metropolitan Atlanta, 1985 through 1987

	White			Black			Total		
	n	Prevalence	95% CI	n	Prevalence	95% CI	n	Prevalence	95% CI
Male	33	1.1	0.8, 1.6	24	1.4	0.9, 2.1	57	1.2	0.9, 1.6
Female	24	0.9	0.6, 1.4	19	1.2	0.7, 1.8	43	1.0	0.7, 1.3
Total	57	1.0	0.8, 1.3	43	1.3	0.9, 1.7	100	1.1	0.9, 1.4

Note. CI = confidence interval.

TABLE 3—Age of Infection among 10-Year-Old Children Who Are Hearing-Impaired as a Result of Meningitis, Metropolitan Atlanta, 1985 through 1987

	Haemophilus influenzae Meningitis		Other Meningitis	
Age of Infection	No.	%	No.	%
0 to < 12 mo	8	61.5	4	57.1
12 to < 24 mo	4	30.8	1	14.3
24 to < 36 mo	1	7.7	1	14.3
≥ 36 mo	0	. . .	1	14.3

other disabilities and etiology. Although we did not include information from private schools, it is unlikely that we missed many cases because the children were enrolled in private schools.[1]

Etiology was defined as the underlying condition that led to hearing loss as documented on records from all ascertainment sources. Age at diagnosis was defined as the child's age the first time a hearing loss was documented in the available records.

Using data provided by the Georgia Office of Planning and Budget, we estimate that 89 534 10-year-old children lived in the study area during the years 1985 through 1987. Race was defined as White or Black because 96% of all 10-year-old children in Atlanta who were not White were Black, and 97% of the children identified as hearing-impaired were either White or Black.

Results

We identified 100 10-year-old hearing-impaired children (1.1 per 1000) (Table 1). The prevalence of hearing loss averaging at least 71 dB was 0.7 per 1000 children (95% confidence interval [CI] = 0.6, 0.9), and the prevalence of hearing loss averaging at least 90 dB was 0.5 per 1000 (95% CI = 0.4, 0.7).

Most of the children (88%) could be identified through the public school systems. Three children were enrolled in public school but were not included on the list provided to us, and nine were enrolled in state-supported schools for the hearing-impaired. Ninety-two percent of the children had received an audiogramwithin 3 years of ascertainment, and all but four cases had assessments after entering school. Overall, 93% of the assessments used to determine study eligiblity were by pure tone audiometry.

The etiology of the hearing loss was unknown for 55% of the children (Table 2). The prevalence of hearing impairment caused by *Haemophilus influenzae* meningitis was 0.2 per 1000 (95% CI = 0.1, 0.3). The rates of hearing impairment caused by other types of meningitis and by congenital rubella syndrome were each 0.1 per 1000 (95% CI = 0.03, 0.2).

Many cases of hearing impairment could be attributed to events, such as genetic conditions and congenital rubella syndrome, that occurred prenatally. In addition, *H influenzae* and other types of meningitis leading to hearing loss often occurred during infancy (Table 3). Only one case of hearing impairment was attributed to meningitis occurring in a child older than 36 months.

TABLE 2—Etiology of Hearing Impairment among 100 Hearing-Impaired 10-Year-Old Children Living in Metropolitan Atlanta, 1985 through 1987

	No.
Unknown	55
Genetic/hereditary[a]	13
Haemophilus influenzae meningitis	13
Other postnatal meningitis	7
Congenital rubella syndrome	7
Ototoxic drug	3
Cochlear dysplasia	1
Mastoiditis	1

[a]Includes one child with hearing impairment attributed to Soto syndrome and two children with hearing impairment attributed to neurofibromatosis. The other 10 cases were attributed to a genetic condition on the basis of a family history.

Although hearing impairment frequently developed before age 2, only 18 of the hearing-impaired children were known to have been diagnosed before the age of 2 years. Four of these were children in whom hearing impairment was attributed to a prenatal or genetic event, 1 was a child exposed to ototoxic drugs, 8 were children for whom the etiology of the hearing loss was unknown, and 5 were children who had had meningitis before the age of 2. Nearly half (n = 46) of the children were diagnosed before the age of 4 years and two thirds (n = 67) were diagnosed before the age of 6 years.

Nearly three quarters (n = 72) of the children with a hearing impairment did not have any of the other disabilities ascertained in the Metropolitan Atlanta Developmental Disabilities study (Table 4). Twenty-one hearing-impaired children also had mental retardation (0.2 per 1000, 95% CI = 0.1, 0.4; 10 children had mild mental retardation and 11 were severely retarded); 10 had epilepsy (0.1 per 1000, 95% CI = 0.05, 0.2); 9 had cerebral palsy (0.1 per 1000, 95% CI = 0.05, 0.2); and 3 were legally blind (0.03 per 1000, 95% CI = 0.01, 0.10).

Discussion

We observed a prevalence of childhood hearing impairment (1.1 per 1000 children) that is comparable to previously reported rates for developed countries (0.56 to 2.6 per 1000).[2-5] Ascertainment

TABLE 4—Number of Hearing-Impaired 10-Year-Old Children with Other Selected Disabilities, Metropolitan Atlanta, 1985 through 1987

Other Disabilities	No.
None	72
Mental retardation only	13
Epilepsy only	3
Cerebral palsy only	1
Blindness only	1
Mental retardation and cerebral palsy	1
Mental retardation and blindness	2
Cerebral palsy and epilepsy	2
Mental retardation, cerebral palsy, and epilepsy	5

methods and diagnostic differences may explain some of the variation in rates. We found that 57% of the children with a hearing impairment in Atlanta were boys, a proportion that is similar to that reported by other researchers.[2,4,6,7] The prevalence of hearing impairment was slightly higher among Black children than among White children; this finding needs confirmation in other studies.

In the present analysis, more than one quarter (n = 28) of the children with a hearing impairment had at least one of four other developmental disabilities. Other researchers have reported that 20% to 43% of children with a hearing impairment have another medical condition or disability.[4,7-9] It is difficult, however, to compare these proportions because of differences in the other conditions included and in ascertainment methods.

The prevalence of hearing impairment caused by congenital rubella infection has been reported to be as high as 0.5 per 1000 in children born before development of the rubella vaccine in 1970.[5,8,10] In contrast, we found the prevalence of hearing impairment resulting from congenital rubella syndrome to be 0.1 per 1000. The occurrence of hearing impairment resulting from this syndrome is likely to have decreased further among children born since 1977.[11,12]

Meningitis caused 20% of the cases of childhood hearing loss in this study. Previous researchers found that meningitis caused 3% to 13% of all childhood hearing impairments.[4,5,7,9,13] This difference may reflect changes in the occurrence of meningitis or hearing impairment or changes in the effectiveness of treatment for meningitis. Our findings confirm an earlier report that most cases of meningitis that lead to hearing impairments occur early in life,[7] and they correspond to what is known about the timing of *H influenzae* infections.[14] Because the incidence of *H influenzae* infection has already been reduced by the use of a vaccine,[15] the prevalence of childhood hearing impairment may be reduced by vaccinating infants as young as 2 months for *H influenzae*.[14,16]

The etiology of the hearing impairment could not be determined for a larger proportion of children in this study than has been reported in most other studies (i.e., 25%–48%).[5,7,8,13,17-19] This difference could reflect limitations in our method for determining etiology or differences in the definition of what constitutes a known etiology.

As has been found in previous studies, a large proportion of our cases of childhood hearing impairment, even those that developed prenatally or in infancy, were not diagnosed within the first 2 years of the child's life.[7,20-22] This finding does not appear to be an artifact of our ascertainment methods or of our inclusion of children born outside the study area, since the finding was not substantially altered when we limited our analysis to children born in the study area. These results are disturbing because of the importance of early intervention in the development of social skills and language.

Our data suggest that the development of rubella vaccination programs has already reduced the occurrence of hearing impairment and that vaccinating young infants against *H influenzae* may further reduce its occurrence. However, our data also suggest that many cases of childhood hearing impairment are not preventable, given current knowledge, because the etiology is either genetic or unknown. Thus, efforts need to be directed toward earlier identification and intervention. □

Acknowledgments

This report was supported in part by funds from the Comprehensive Environmental Response, Compensation, and Liability Act trust funds through an interagency agreement with the Agency for Toxic Substances and Disease Registry, Public Health Service, US Department of Health and Human Services.

References

1. Yeargin-Allsopp M, Murphy CC, Oakley GP, et al. A multiple-source method for studying the prevalence of developmental disabilities in children: the Metropolitan Atlanta Developmental Disabilities Study. *Pediatrics* 1992;89:624–630. (Published erratum appears in *Pediatrics*, 1992;90:1001.)
2. Schein JD. Childhood hearing loss: epidemiology and implications. *Birth Defects.* 1980;16:3–7.
3. Upfold LJ, Isepy J. Childhood deafness in Australia: incidence and maternal rubella, 1949–1980. *Med J Aust.* 1982;2:323–326.
4. Martin JAM. Aetiological factors relating to childhood deafness in the European Community. *Audiology.* 1982;21:149–158.
5. Davidson J, Hyde ML, Alberti PW. Epidemiology of hearing impairment in childhood. In: Barajas JJ, Borg E, eds. *Overview and Evaluation of the Hearing Impaired Child.* Stockholm, Sweden: Almquist & Wiskele Periodical Co; 1987:13–20.
6. Trybus RJ, Karchmer MA, Kerstetter PP, Hicks W. The demographics of deafness resulting from maternal rubella. *Am Ann Deaf.* 1980;125:977–984.
7. Das VK. Aetiology of bilateral sensorineural deafness in children. *J Laryngol Otol.* 1988;102:975–980.
8. Budden SS, Robinson GC, MacLean CD, Cambon KG. Deafness in infants and preschool children. *Am Ann Deaf.* 1974;119:387–395.
9. Schildroth A. Hearing-impaired children under age 6: 1977 and 1984. *Am Ann Deaf.* 1986;131:85–90.
10. Upfold LJ. Changes in the significance of maternal rubella as a factor in childhood deafness—1954 to 1982. *Med J Aust.* 1984;140:641–644.
11. Menser MA, Hudson JR, Murphy AM, Cossart YE. Impact of rubella vaccination in Australia. *Lancet.* 1984;1:1059–1061.
12. Upfold LJ. Rubella-related deafness. *Med J Aust.* 1988;149:723. Letter.
13. Schildroth A, Terrero I, Hotto S, Lam K. Hearing-impaired children in Venezuela: 1985. *Int J Rehab Res.* 1987;10:185–195.
14. *H. influenzae* vaccine for infants. *Med Lett Drugs Ther.* 1991;33:5–7.
15. Adams WG, Deaver KA, Cochi SL, et al. Decline of childhood *Haemophilus influenzae* type b (Hib) disease in the Hib vaccine era. *JAMA.* 1993;260:321–325.
16. American Academy of Pediatrics Committee on Infectious Diseases. *Haemophilus influenzae* type b conjugate vaccines: recommendations for immunization of infants and children 2 months of age and older: update. *Pediatrics.* 1991;88:169–172.
17. Feinmesser M, Tell L, Levi H. Etiology of childhood deafness with reference to the group of unknown cause. *Audiology.* 1986;25(2):65–69.
18. Lenzi A, Zaghis A. Incidence of genetic factors in the causation of deafness in childhood. *Scand Audiol Suppl.* 1988;30:37–41.
19. Pabla HS, McCormick B, Gibbin KP. Retrospective study of the prevalence of bilateral sensorineural deafness in childhood. *Int J Pediatr Otorhinolaryngol.* 1991;22:161–165.
20. Upfold LJ. The early detection of childhood deafness. *Med J Aust.* 1978;2:507–510.
21. Lyon DJ, Lyon ME. Early detection of hearing loss. *Can J Public Health.* 1982;73:410–415.
22. Pappas DG. A study of the high risk registry for sensorineural hearing impairment. *Otolaryngol Head Neck Surg.* 1983;91:41–44.

Socioeconomic Factors and Height of Preschool Children in the Czech Republic

Martin Bobák, MD, MSc, Bohumír Kříž, MD, PhD, David A. Leon, PhD, Jana Dáňová, MD, and Michael Marmot, FFPHM, PhD

ABSTRACT

The effect of socioeconomic factors on growth was investigated among 2275 children 3 to 6 years old attending nurseries in the Czech Republic. Measured heights of children were converted into height-for-age *z* scores. After adjustment for birthweight, parental height, and other socioeconomic variables, only mother's education was independently and significantly associated with children's height; adjusted differences in *z* scores between children of secondary- and university-educated mothers and children of mothers with only primary education were 0.12 and 0.31, respectively. This is equivalent to 0.5 and 1.5 cm, respectively, for children 5 years old. (*Am J Public Health.* 1994;84:1167–1170)

Introduction

Growth in childhood is considered to be a sensitive indicator of children's health.[1] It is susceptible to environmental factors and is thus useful for monitoring the effect of such factors on children.[2–6] Height in childhood is a good predictor of height in adulthood,[7] and adult height has been associated with mortality from a range of causes.[6,8]

Differences in children's height, which persist into adulthood,[9] have been reported between groups characterized by a wide range of socioeconomic factors such as social class based on parental occupation,[10–13] employment status,[5] family size and birth order,[3,5,10,12,14] parental education,[5,13] parental occupation,[5] and homelessness.[15] Because of the close correlation between education and other socioeconomic factors, previous attempts to estimate the independent effect of education have been inconclusive.[12,16]

Little is known of the magnitude of the effect of socioeconomic circumstances on health in the former socialist countries of central and eastern Europe. This analysis was conducted with three objectives: (1) to explore whether socioeconomic differences exist in growth impairment in the Czech Republic, (2) to investigate whether these differences are independent of parental height and birthweight, and (3) to determine which of the socioeconomic factors is most important.

Subjects and Methods

Preschool children 3 to 6 years of age attending 21 nurseries in five towns of the Czech Republic (Beroun, Most, Prague 3 district, Pribram, and Strakonice) were studied. All children registered in the nurseries in October 1992 were considered to be eligible for the study.

Parents were asked to complete a questionnaire on socioeconomic circumstances and life-style of the family, parents' height and weight, and children's health in the previous 12 months. Mothers and fathers were classified into three educational levels: primary, secondary, and university. A binary variable on crowding was calculated; families with more than one person per room were considered as living in crowded conditions. Information was also collected on car and weekend house ownership, whether the child lived with both parents, how often the child spent the weekend in the countryside, family size, birth order, birthweight, frequency of respiratory symptoms in the previous 12 months, maternal smoking (more than one cigarette a day), and history of breast-feeding.

Children's height and weight were measured in nurseries by centrally trained local public health workers (one to two in each town). A steel stadiometer, combined with a weighing device, was used in a standard fashion. Height was recorded to the nearest half centimeter. Because the height of young children depends critically on age, deviation from a growth standard was used instead of absolute measures. Epi-Info's anthropometric module[17] was used to convert individual anthropometric measurements to height-for-age *z* scores based on the US growth standards of the National Center for Health Statistics.[18] The height-for-age *z* score expresses, in terms of standard deviations, the extent to which the height of individual children differs from the value predicted by the growth standard

Martin Bobák, Bohumír Kříž, and Jana Dáňová are with the Department of Epidemiology, Charles University, Prague, Czech Republic. Martin Bobák and David A. Leon are with the Department of Epidemiology and Population Sciences, London School of Hygiene and Tropical Medicine. Bohumír Kříž is also with the Centre of Epidemiology and Microbiology, National Institute of Public Health, Prague. Michael Marmot is with the Department of Epidemiology and Public Health, University College London.

Requests for reprints should be sent to Martin Bobák, MD, MSc, Department of Epidemiology and Population Sciences, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, United Kingdom.

This paper was accepted January 10, 1994.

TABLE 1—Descriptive Characteristics of Children in the Study

	Sample (n = 2275)
Age, mo, mean (SD)	57.5 (11.8)
Birthweight, g, mean (SD)	3338.7 (530.8)
Height-for-age *z* score, mean (SD)	0.49 (1.13)
Maternal age, y, mean (SD)	29.7 (4.9)
Sex, no. (%)	
Male	1177 (51.7)
Female	1098 (48.3)
Mother's education, no. (%)	
Primary	902 (39.6)
Secondary	1038 (45.6)
University	327 (14.4)
Missing	8 (0.4)
Father's education, no. (%)	
Primary	1120 (49.2)
Secondary	638 (28.0)
University	398 (17.5)
Missing	119 (5.2)
Car ownership, no. (%)	
No	796 (35.0)
Yes	1477 (64.9)
Missing	2 (0.1)
Siblings, no. (%)	
None	712 (31.3)
1	1258 (55.3)
2	257 (11.3)
3	40 (1.8)
4 or more	7 (0.3)
Missing	1 (0.0)
Crowding, no. (%)	
No	1434 (63.0)
Yes	818 (36.0)
Missing	23 (1.0)
Weekend house ownership, no. (%)	
No	1334 (58.6)
Yes	937 (41.2)
Missing	4 (0.2)
Breast-feeding, no. (%)	
No	142 (6.2)
Yes	2136 (93.7)
Missing	2 (0.1)
Maternal smoking, no. (%)	
No	1417 (62.1)
Yes	852 (37.4)
Missing	11 (0.5)

for their age and sex. (In absolute terms, one standard deviation represents approximately 4 to 5 cm in our data.)

TABLE 2—Differences in Height-for-Age *z* Score in Preschool Children in Five Czech Towns

	Crude			Adjusted for Biological Factors[a]			Fully Adjusted[b]		
	Difference	SE	*P*	Difference	SE	*P*	Difference	SE	*P*
Mother's education									
Primary	0.00	...	...	0.00	...	...	0.00	...	...
Secondary	0.27	0.05	<.001	0.11	0.05	.019	0.12	0.06	.030
University	0.46	0.07	<.001	0.27	0.07	<.001	0.31	0.09	<.001
Trend			<.001			<.001			<.001
Father's education									
Primary	0.00	...	...	0.00	...	...	0.00	...	...
Secondary	0.14	0.06	.012	0.05	0.05	.322	−0.01	0.06	.908
University	0.19	0.07	.003	0.08	0.06	.176	−0.08	0.08	.308
Trend			<.001			.154			.428
No. of siblings									
None	0.00	...	...	0.00	...	...	0.00	...	...
1	−0.14	0.05	.011	−0.10	0.05	.042	−0.09	0.06	.100
2	−0.28	0.08	<.001	−0.09	0.08	.215	−0.03	0.09	.746
3	−0.53	0.18	.003	−0.20	0.17	.228	0.01	0.21	.980
4 or more	−0.61	0.42	.148	−0.28	0.38	.462	−0.27	0.41	.515
Trend			<.001			.047			.426
Car ownership									
No	0.00	...	...	0.00	...	...	0.00	...	...
Yes	0.14	0.05	.006	0.08	0.05	.090	0.08	0.05	.146
Crowding									
No	0.00	...	...	0.00	...	...	0.00	...	...
Yes	−0.16	0.05	.001	−0.07	0.05	.129	−0.01	0.06	.910
Weekend house									
No	0.00	...	...	0.00	...	...	0.00	...	...
Yes	0.06	0.05	.210	−0.02	0.04	.632	−0.02	0.05	.716
Birthweight (per 1000 g)	0.45	0.04	<.001	0.29	0.04	<.001	0.34	0.04	<.001
Maternal height (per 10 cm)	0.64	0.04	<.001	0.47	0.04	<.001	0.46	0.03	<.001
Paternal height (per 10 cm)	0.56	0.03	<.001	0.40	0.03	<.001	0.39	0.04	<.001
Breast-feeding									
No	0.00	...	...	0.00	...	...	0.00	...	...
Yes	0.04	0.10	.707	−0.07	0.09	.446	0.01	0.10	.918
Maternal smoking									
No	0.00	...	...	0.00	...	...	0.00	...	...
Yes	−0.13	0.05	.007	−0.09	0.05	.040	−0.02	0.05	.650
Respiratory symptoms									
No	0.00	...	...	0.00	...	...	0.00	...	...
Yes	−0.09	0.06	.115	−0.08	0.06	.164	−0.06	0.06	.298

[a]Adjusted for birthweight and height of mother and father.
[b]Adjusted for biological and all socioeconomic variables.

The height-for-age *z* score was normally distributed. Differences in *z* scores between levels of explanatory variables were estimated with normal errors linear regression[19]; *z* scores, birthweight, and height of parents were modeled as continuous variables. Differences in *z* scores were adjusted first for "biological factors" (mother's and father's self-reported height and birthweight reported by the mother) and then for socioeconomic factors.

Results

Of the 2979 children registered in nurseries, 143 had never attended the nursery. Completed questionnaires were returned for 2366 children, and valid height was available for 2275 children (76% of those registered).

Table 1 describes the characteristics of our study population. The Czech children were almost half a standard

deviation taller than the American standards, the mean height-for-age z score being 0.49. Relatively more mothers had completed secondary school than fathers. About two thirds of families owned a car, and only 14% of children had two or more siblings. There was a positive association between mother's and father's education (data not shown).

Of the socioeconomic variables, education of parents, car ownership, crowding, number of siblings, and maternal smoking were significantly associated with height-for-age z score (see Table 2). Among the other variables, mother's and father's height and birthweight were most strongly related to z score; the multiple R^2 for these three variables was .20 (all P values less than .0001).

The effects of the various explanatory factors on height-for-age z score, adjusted for biological factors and then for all factors simultaneously, are shown in Table 2. After adjustment for birthweight and height of parents, maternal education and number of siblings retained their significant association with height-for-age z score. When all socioeconomic variables plus birthweight and parental height were included in one model, maternal education was the only socioeconomic variable that remained significantly associated with z score; children with primary-educated mothers were 0.12 and 0.31 standard deviations shorter than children with secondary- and university-educated mothers, respectively. This is equivalent to 0.5 and 1.5 cm, respectively, for 5-year-old children.

Discussion

Our study has several limitations. Only about 80% of all children in the Czech Republic attend nurseries,[14] and the five towns that were selected may not be fully representative of the whole country. While this should not affect the internal validity of the study, the effect of nonresponse is more difficult to assess. The cross-sectional design of the study is not optimal for establishing a causal association. All explanatory variables were self-reported, and a certain degree of misclassification cannot be excluded. This is particularly pertinent with respect to birthweight and height of mother and father. However, birthweight reported by mothers seems to agree well with medical records,[12] and, although adults tend to overestimate their height in a nonrandom fashion,[12,20,21] this bias is not independently related to social position.[21] Misclassification of this sort is unlikely to account for the strong association between maternal education and children's height.

Apart from birthweight and parental height, only maternal education was significantly related to children's growth in multiple regression, while the effect of other socioeconomic variables reflecting material conditions (crowding, car and weekend house ownership) disappeared.

The decline in the effect of maternal education after adjustment for parental height and birthweight suggests that some of the effect may be mediated through genetic and/or intrauterine factors. However, some of the effect must also be due to factors operating in childhood, because there remains a substantial difference in height by maternal education after adjustment for birthweight and parental height.

The mechanism in childhood through which maternal education affects growth is not clear. Our results do not support suggestions that breast-feeding, current maternal smoking,[22] or frequent disease[6,23] (Table 2) are related to height of children. Nutrition, however, is likely to play an important mediating role,[3,4,6,24,25] although only a few studies have investigated this question in developed countries.[26,27] Unfortunately, little is known about differences in nutrition by socioeconomic groups in the Czech population, although the expected differences would be reduced by the fact that children have their meals in nurseries. It has been suggested that maternal care and the nature of the mother–child relationship can affect growth.[26] However, our data shed no light on this issue. Further work is required to explain the central role of maternal education as distinct from other socioeconomic factors. □

Acknowledgments

This study was supported by grants from the Czech Ministries of Health and Environment. Martin Bobák is a Wellcome Trust Fellow in Clinical Epidemiology.

References

1. Waterlow JC, Buzina R, Keller W, Lane M, Nichaman MZ, Tanner JM. The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. *Bull WHO.* 1977;55:489–498.
2. Rona RJ. Genetic and environmental factors in the control of growth in childhood. *Br Med Bull.* 1981;37:265–272.
3. Tanner JM. *Foetus into Man. Physical Growth from Conception to Maturity.* London, England: Open Books; 1978.
4. Department of Health and Social Security. *Reports on Health and Social Subjects No. 6. First Report by the Sub-committee on Nutritional Surveillance.* London, England: Her Majesty's Stationery Office; 1973.
5. Rona RJ, Wainwright AH, Altman DG, Irwig LM, Florey CV. Surveillance of growth as a measurement of health in the community. In: Holland WW, Ipsen J, Kostrzewski J, eds. *Measurement of Levels of Health.* Copenhagen, Denmark: WHO Regional Office for Europe; 1979:397–404.
6. Floud R, Wachter K, Gregory A. *Height, Health and History: Nutritional Status in the United Kingdom, 1750–1980.* Cambridge, England: Cambridge University Press; 1990.
7. Tanner JM, Whitehouse RH, Marshall WA, Carter BS. Prediction of adult height from height, bone age, and occurrence of menarche, at ages 4 to 16 with allowance for midparent height. *Arch Dis Child.* 1975;50:14–26
8. Marmot MG, Shipley MJ, Rose G. Inequalities in death: specific explanations of a general pattern? *Lancet.* 1984;i:1003–1006.
9. Kuh D, Wadsworth M. Parental height: childhood environment and subsequent adult height in a national birth cohort. *Int J Epidemiol.* 1989;18:663–668.
10. Goldstein H. Factors influencing the height of seven-year-old children: results from the National Child Development Study. *Hum Biol.* 1971;43:92–111.
11. Smith AM, Chinn S, Rona RJ. Social factors and height gain of primary schoolchildren in England and Scotland. *Ann Hum Biol.* 1980;7:115–124.
12. Rona RJ, Swan AV, Altman DG. Social factors and heignt of primary schoolchildren in England and Scotland. *J Epidemiol Community Health.* 1978;32:147–154.
13. Rona RJ, Chinn S. Height of children from ethnic groups in England. *Ann Hum Biol.* 1986;13:453–471.
14. Bouchalova M. *Vyvoj behem detstvi a jeho ovlivneni* (Development in childhood and its determinants). Prague, Czechoslovakia: Avicenum; 1987.
15. Fierman AH, Dreyer BP, Quinn L, et al. Growth delay in homeless children. *Pediatrics.* 1991;88:918–925.
16. Bielicki T. Physical growth as a measure of the economic well-being of populations: the twentieth century. In: Falkner F, Tanner JM, eds. *Human Growth. A Comprehensive Treatise. Volume 3. Methodology, Ecological, Genetic, and Nutritional Effects on Growth.* 2nd ed. New York, NY: Plenum Press, 1986:283–305.
17. Dean AD, Dean JA, Burton JH, Dicker RC. *Epi Info, Version 5: A Word Processing, Database, and Statistics Program for Epidemiology on Microcomputers.* Atlanta, Ga: Centers for Disease Control; 1990.
18. Dibley MJ, Goldsby JB, Staehling NW, Trowbridge FL. Development of normalized curves for the international growth reference: historical and technical considerations. *Am J Clin Nutr.* 1987;46:736–748.
19. *SAS/STAT User's Guide. Release 6.03 Edition.* Cary, NC: SAS Institute Inc; 1988.
20. Stewart AL. The reliability and validity of self-reported weight and height. *J Chronic Dis.* 1982;35:295–309.

21. Palta M, Prineas RJ, Berman R, Hannan P. Comparison of self-reported and measured height and weight. *Am J Epidemiol.* 1982;115:223–230.
22. Rona RJ, Chinn S, Florey CV. Exposure to cigarette smoking and children's growth. *Int J Epidemiol.* 1985;14:402–409.
23. Rona RJ, Florey CV. National Study of Health and Growth: respiratory symptoms and height in primary schoolchildren. *Int J Epidemiol.* 1979;8:35–43.
24. Zerfas AJ, Jelliffe DB, Jelliffe EF. Epidemiology and nutrition. In: Falkner F, Tanner JM, eds. *Human Growth. A Comprehensive Treatise. Volume 3. Methodology, Ecological, Genetic, and Nutritional Effects on Growth.* 2nd ed. New York, NY: Plenum Press; 1986:475–499.
25. Growth and development. In: Keys A, Brozek J, Henschel A, Mickelsen O, Taylor HL, eds. *The Biology of Human Starvation.* Minneapolis, Minn: University of Minnesota Press; 1950;2:974–1001.
26. Pollitt E. Failure to thrive: socioeconomic dietary intake and mother-child interaction data. *Fed Proc.* 1975;34:1593–1597.
27. Sub-committee on Nutritional Surveillance, Committee on Medical Aspects of Food Policy. *The Diets of British Schoolchildren.* London, England: Her Majesty's Stationery Office; 1989.

A2623-1

American Journal of Public Health

August 1994, Vol. 84, No. 8 Established 1911

Editorials

1211 Water Chlorination, Mutagenicity, and Cancer Epidemiology *K. P. Cantor*
1213 Making Aging a Public Health Priority *R. Kane*
1214 Changing Lives and Changing Social Structures: Common Concerns of Social Science and Public Health *M. W. Riley*

Articles

VARIOUS TOPICS

1218 The Demographic Distribution of US Drinking Patterns in 1990: Description and Trends from 1984 *L. T. Mulanik and W. B. Clark*
1223 Drinking Water Mutagenicity and Gastrointestinal and Urinary Tract Cancers: An Ecological Study in Finland *M. Koivusalo, J. J. K. Jaakkola, T. Vartiainen, T. Hakulinen, S. Karjalainen, E. Pukkala, and J. Tuomisto*
1229 A National Study of Transitional Hospital Services in Mental Health *R. A. Dorwart and C. W. Hoover*
1235 What Determines Mortality Risk in Male Former Cigarette Smokers? *Y. Ben-Shlomo, G. D. Smith, M. J. Shipley, and M. G. Marmot*
1243 The Risk of HIV Infection in a National Sample of Women with Injection Drug–Using Partners *S. Tortu, M. Beardsley, S. Deren, and W. R. Davis*
1250 A Recent Decline in Cocaine Use among Youthful Arrestees in Manhattan, 1987 through 1993 *A. Golub and B. D. Johnson*

AGING

1255 Cognitive Impairment and Mortality in Older Community Residents *H. R. Kelman, C. Thomas, G. J. Kennedy, and J. Cheng*
1261 The US Economic and Social Costs of Alzheimer's Disease Revisited *R. L. Ernst and J. W. Hay*
1265 The Burden and Outcomes Associated with Dehydration among US Elderly, 1991 *J. L. Warren, W. E. Bacon, T. Harris, A. M. McBean, D. J. Foley, and C. Phillips*
1270 The Groningen Activity Restriction Scale for Measuring Disability: Its Utility in International Comparisons *T. P. B. M. Suurmeijer, D. M. Doeglas, T. Moum, S. Briançon, B. Krol, R. Sanderman, F. Guillemin, A. Bjelle, and W. J. A. van den Heuvel*
1274 Functional Transitions among the Elderly: Patterns, Predictors, and Related Hospital Use *V. Mor, V. Wilcox, W. Rakowski, and J. Hiris*
1281 The Outcomes of Patients Newly Admitted to Nursing Homes after Hip Fracture *D. P. Kiel, A. Eichorn, O. Intrator, R. A. Silliman, and V. Mor*
1287 Treatment and Survival among Elderly Americans with Hip Fractures: A Population-Based Study *G. L. Lu-Yao, J. A. Baron, J. A. Barrett, and E. S. Fisher*
1292 Lumbar Spine Surgery and Mortality among Medicare Beneficiaries, 1986 *N. B. Oldridge, Z. Yuan, J. E. Stoll, and A. R. Rimm*
1299 Diet, Indicators of Kidney Disease, and Later Mortality among Older Persons in the NHANES I Epidemiologic Follow-Up Study *J. T. Dwyer, J. H. Madans, B. Turnbull, J. Cornoni-Huntley, C. Dresser, D. F. Everett, and R. D. Perrone*

Public Health Briefs

VARIOUS TOPICS

1304 Emotional Disability Days: Prevalence and Predictors *A. C. Kouzis and W. W. Eaton*

(continued on page 1201)

American Public Health Association

Featuring Aging

Annual Meeting Preliminary Program

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

August 1994, Vol. 84, No. 8 ISSN: 0090-0036 Established 1911

1307 The Prevalence and Demographic Predictors of Illicit and Licit Drug Use among Lesbians and Gay Men *W. F. Skinner*

1310 Breast Biopsy Rate and Positivity in Rhode Island *M. S. Richards, J. Feldman, R. A. Smith, and B. A. DeBuono*

AGING

1313 Projected Numbers of Cancers Diagnosed in the US Elderly Population, 1990 through 2030 *A. P. Polednak*

1316 Subsequent Hip Fracture among Older Adults *F. D. Wolinsky and J. F. Fitzgerald*

1319 Lifetime Milk Consumption and Bone Mineral Density in Older Women *S. Sorensen, T. L. Holbrook, S. Edelstein, and E. Barrett-Connor*

1322 Attitudes toward Body Size and Dieting: Differences between Elderly Black and White Women *J. Stevens, S. K. Kumanyika, and J. E. Keil*

1325 The 1966 Enactment of Medicare: Its Effect on Discharges from Los Angeles County–Operated Hospitals *P. A. Glassman, R. M. Bell, and R. E. Tranquada*

1327 Sociodemographic and Health Factors in Driving Patterns after 50 Years of Age *R. Kington, D. Reuben, J. Rogowski, and L. Lillard*

Health Law and Ethics

1330 Patients' Rights to Care under Clinton's Health Security Act: The Structure of Reform *W. K. Mariner*

Notes from the Field

1336 Grassroots Efforts to Promote Tobacco-Free Schools in Rural Missouri *M. H. Eschen, R. C. Brownson, J. R. Davis, L. R. Cooperstock, R. Crawford, D. Freeman, G. Howard, and M. J. Michael*

1337 C.A.R.E. Fair: One-Stop Shopping for an Underserved Population *R. Melland, B. Mauntel, and P. Contreras*

1338 A Lead-Safe Family Shelter in an Urban Minority Community *M. R. Farfel and R. Quinn*

Other Journal Departments

1202 APHA Masthead

1208 What *AJPH* Authors Should Know

1209 Advertisers' Index

1211 Journal Masthead

1340 Letters to the Editor

1343 Book Corner

1346 Association News: Obituaries and Forty-Year Members

1352 Job Opportunities

Preliminary Program of the 1994 APHA Annual Meeting

See page P-2 for contents

American Public Health Association

American Journal of Public Health

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 15th St, NW, Washington, DC 20005; (202) 789-5600; TDD (202) 789-5673. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*
Helen Rodriguez-Trias, MD, *Immediate Past President*
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Leonard M. Schuman, MD, *Vice-President (USA)*
John E. F. Hastings, MD, MPH, *Vice-President (Canada)*
Oscar Feo, MD, MSc, *Vice-President (Latin America)*
Jay H. Glasser, PhD, *Treasurer*
Allan G. Rosenfield, MD, *Chair, Executive Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*
Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*
Carol Easley Allan, PhD, RN (1997)
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Patricia P. Evans, MPH (1994)
Eugene Feingold, PhD, JD, *President*
Jay H. Glasser, PhD, *Treasurer*
Alan R. Hinman, MD, MPH (1995)
Margo D. Kerrigan, MPH (1995)
Joanne Lukomnik, MD (1996), *Vice-Chair*
Steven Uranga McKane, DMD, MPH (1995)
Carlos W. Molina, EdD (1997)
Linda Rae Murray, MD, MPH (1996)
Hila Richardson, DrPH, MPH (1996)
Helen Rodriguez-Trias, MD, *Immediate Past President*
Reed V. Tuckson, MD (1994)
Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*
Frances Kissling, *Chair, Action Board*
George E. Pickett, MD, MPH, *Speaker, Program Development Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*

The *American Journal of Public Health* (*AJPH*), published monthly, is the official Journal of the American Public Health Association, Inc, 1015 15th St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005. *AJPH* allotted $40; *The Nation's Health* allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters. Printed in the United States.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on the first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, *AJPH* Reprints, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the Journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The *American Journal of Public Health* is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

The Demographic Distribution of US Drinking Patterns in 1990: Description and Trends from 1984

Lorraine T. Midanik, PhD, and Walter B. Clark, MA

ABSTRACT

Objectives. Since 1981, per capita consumption of alcohol (based on sales figures) has decreased in the United States. This study describes drinking patterns in the 1990 national alcohol survey by demographic correlates and assesses changes in drinking patterns from the 1984 survey.

Methods. Data were obtained from a national household probability sample within the 48 contiguous states; face-to-face interviews were conducted with 2058 adults. The instrument contained questions pertaining to the respondent's background, attitudes toward alcohol, and use of alcohol.

Results. The proportions of current drinkers; current drinkers of wine, beer, and liquor; weekly drinkers; and drinkers who reported having five or more drinks per occasion at least once a week were significantly lower in 1990 than in 1984. These changes remained significant when demographic characteristics were controlled by logistic regression. The findings held for Whites only; there were no significant trends for Blacks or Hispanics.

Conclusions. While there has been a downward turn in alcohol use in the United States, the correlates of alcohol use have not changed. How these shifts affect alcohol-related problems is an important area for future research. (*Am J Public Health.* 1994;84:1218–1222)

Introduction

National alcohol-use surveys in the United States, beginning with the Gallup poll in 1939, originally aimed at differentiating drinkers from nondrinkers. Later surveys focused on a wide array of drinking practices within the US household population. From these data, researchers have examined correlates of different patterns of drinking and assessed trends within certain subgroups.[1]

With the rise in per capita consumption indicated by sales data in the 1960s and 1970s, national surveys became important ways to monitor whether these increases were uniform throughout the population or within specific segments. Results were mixed. Johnson et al.[2] found in their analysis of seven national studies conducted from 1971 through 1976 that there was an overall increase in the proportion of heavier-drinking men and moderate-drinking women aged 35 through 49 years. Little change was detected in rates of abstention and rates of light, moderate, or heavy drinking in the US population in 1979,[3] and the 1984 national alcohol survey found few significant differences between 1967 and 1984.[4] An exception was a small increase in the number of male abstainers. Recently, Williams and Debakey[5] reported an increase in abstention and a decrease in heavier drinking between 1983 and 1988, based on National Health Interview Survey data.

Alcohol sales data through 1990 indicate that per capita consumption of alcohol fell from a high of 2.76 gallons per person aged 14 years or older in 1981 to 2.46 gallons in 1990—a change of nearly 11%. Liquor consumption fell from 1.02 to 0.78 gallons per person from 1981 to 1990, beer consumption fell from 1.39 to 1.34 gallons per person, and wine consumption fell from 0.35 to 0.33 gallons per person.[6] There are regional differences in apparent consumption of alcoholic beverages, but all census regions share in the downward trend in per capita consumption. However, useful as they are, per capita estimates obscure variations in drinking patterns and their correlates.

The purpose of this study is threefold. First, we will describe US drinking patterns in 1990 by demographic correlates. Second, we will compare the 1990 data on drinking patterns with comparable data from the 1984 survey to assess any changes in drinking patterns. Third, we will use multivariate analysis to determine whether the year of the survey is a significant predictor for any change in drinking patterns from 1984 to 1990 when demographic characteristics are controlled in the model.

Methods

Study Sample

The data were collected by the Institute for Survey Research of Temple University from a probability sample of 5970 housing units in 100 primary sampling units within the 48 contiguous states. Interviews were conducted from January through June 1990 with 2058 adults aged 18 years and older (70% response rate). The average length of an interview was 75 minutes, and respondents were paid $10.

The authors are with the Alcohol Research Group, California Pacific Medical Center Research Institute of San Francisco, Berkeley, Calif. Lorraine T. Midanik is also with the School of Social Welfare, University of California at Berkeley.

Requests for reprints should be sent to Lorraine T. Midanik, PhD, Alcohol Research Group, 2000 Hearst Ave, Berkeley, CA 94709.

This paper was accepted November 18, 1993.

The 1984 national alcohol survey oversampled Black and Hispanic respondents within the 100 primary sampling units, and an additional 10 sampling units were selected from geographical areas with high proportions of Black and Hispanic residents.[1]

Weights were applied to both data sets to make them representative of the national household population in the United States during 1984 and 1990. First, a weight took into account demographic differences (age, sex, and region) and nonresponse rates. Second, an average design-effects weight based on selected alcohol use, alcohol-related problems, and demographic variables was derived, which takes into account clustering effects.[7] For the 1990 survey, the average design-effects weight was 1.79, which reduced the effective sample size from 2058 to 1150. In 1984, the average design-effects weight was considerably higher (4.24) because of the oversampling of Black and Hispanic respondents. Thus, the effective sample size was reduced from 5221 to 1228.

Dependent Variables

The description of patterns of alcohol use is no simple task, as Room's discussion makes clear.[8] Frequency of drinking, beverage type, quantity of alcohol per occasion, overall quantity, and volume per unit of time are all important measures for various purposes.

We used the following dichotomous measures of alcohol use during the 12 months preceding the interview: (1) current drinking (any alcoholic beverage use in the last year); (2) weekly drinking (any alcoholic beverage use at least weekly in the last year); current (3) wine, (4) beer, and (5) liquor drinking (any use of the beverage in the last year); and (6) five or more drinks (drinking five or more drinks on one occasion of any combination of alcoholic beverages weekly or more often during the last year).

The first two measures were based on an overall question about frequency of alcohol use: "How often do you *usually* have *any* kind of beverage *containing alcohol*, whether it is wine, beer, whiskey, or any other drink?" Measures 3 through 5 were based on similar questions about the frequency of use of each beverage type separately. Construction of the sixth measure was complex because there was no direct question that obtained this information. We used two questions, identical on both surveys, to construct the five or more drinks variable: "How often do you usually have wine or wine coolers?" and "Think of all the times you have had wine or wine coolers recently. When you drink wine, how often do you have as many as five or more glasses?" Parallel questions on beer and liquor were asked separately. (Details on the construction of this measure are available from the authors.)

TABLE 1—US Drinking Patterns: Percentages of 1984 and 1990 Survey Respondents Reporting Drinking Behaviors

	1984 (n = 5221)	1990 (n = 2058)	x^2
Current drinking	69.6	65.0	5.54*
Current wine drinking	51.2	43.6	13.95**
Current beer drinking	51.5	45.3	9.35***
Current liquor drinking	51.8	43.5	16.46**
Weekly drinking	35.9	29.0	13.03**
Having 5+ drinks on one occasion at least weekly	6.2	3.9	6.49**

Note. Percentages are based on weighted sample sizes (effective n's of 1228 for the 1984 survey and 1150 for the 1990 survey).
*$P < .05$; **$P < .01$; ***$P < .001$.

Independent Variables

Eleven independent variables were used to describe the subgroups of the population: gender, age, marital status, ethnicity, income, employment status, religion, importance of religion, education, urbanicity, and region.

Analysis

Descriptive comparisons between surveys were made by using a two-tailed difference-of-proportions test for each of the six dependent variables and for comparing proportions between the two surveys within each subgroup of the independent variables. We merged the two data sets and ran logistic regressions on each dichotomous dependent variable to test whether the year of the survey was a significant predictor when the effects of differences in demographic characteristics in these two surveys were controlled.

Results

1990 Cross-Sectional Data

About one third of adults reported no use of alcoholic beverages in the year preceding the survey (Table 1)—although the rate of abstention may be increasing, as will be noted below. Approximately 40% of adults reported drinking each of the three types of beverages in the last year. Just under one third of adults reported current use of any kind of alcoholic beverage as often as once a week, and a small proportion (3.9%) reported that they drank five or more drinks on one occasion at least as often as once a week.

Table 2 presents data on the use of any alcoholic beverage by demographic subgroups. Gender, as usual, was quite strongly associated with alcohol use. A statistically significantly lower proportion of women were current drinkers, weekly drinkers, or consumers of five or more drinks per occasion at least weekly. As in past surveys, age was negatively and significantly ($P < .05$) related to each of the measures of alcohol use and to the use of wine, beer, and distilled beverages considered separately.

Religious affiliation has been consistently associated with alcohol use in the United States. The religious groups here labeled conservative Protestants discourage the use of any alcoholic beverage by their members; other religious groups oppose only excessive use. Because of the small numbers, the comparisons presented here are of conservative Protestants vs all others. Significantly lower proportions of conservative Protestants reported current drinking, weekly drinking, and drinking five more drinks per occasion. Respondents who reported that religion was "very important" in their lives were less likely than others to drink by all measures of use.

We found no statistically significant differences in alcohol use among ethnic groups. This finding is of considerable interest because race and ethnicity have been significantly associated with alcohol use in past surveys, with Whites reporting higher rates of alcohol use on many measures. Whites in the 1990 survey did

TABLE 2—Demographic Characteristics of Drinkers: 1984 and 1990

	1984 (n = 5221)	1990 (n = 2058)	Current Drinkers, %		Weekly Drinkers, %		Drinkers of Five or More Drinks,[a] %	
			1984	1990	1984	1990	1984	1990
Gender								
Male	2093	869	75.9	71.2	48.9	40.0*	10.6	6.5*
Female	3128	1189	63.9	59.4	24.4	18.8*	2.2	1.4
Age, y								
18–29	1515	442	77.8	73.1	40.1	32.2*	10.3	7.0
30–39	1277	520	77.6	70.7	42.3	30.0*	7.7	3.0*
40–49	711	330	68.9	68.4	33.5	29.0	4.6	3.8
50–59	593	228	66.1	61.9	36.9	32.8	4.2	3.3
60+	1092	538	53.5	49.4	25.5	22.4	1.2	1.5
Gender and age								
Male								
18–29 y	621	201	81.7	76.5	52.3	44.4	17.6	11.0
30–39 y	511	209	86.7	72.5*	60.0	39.2*	12.4	4.3*
40–49 y	305	146	77.5	71.8	44.7	38.5	6.0	6.8
50–59 y	235	102	70.7	64.6	48.4	43.5	8.7	6.2
60+ y	413	211	58.8	65.6	36.0	34.8	2.5	3.1
Female								
18–29 y	894	241	73.9	69.7	28.2	19.7	3.1	3.0
30–39 y	766	311	69.3	69.0	26.4	20.9	3.5	1.7
40–49 y	406	184	62.3	65.1	24.8	20.1	3.6	1.0
50–59 y	358	126	62.2	59.8	26.9	24.2	0.2	0.9
60+ y	679	327	48.9	37.0*	16.5	12.9	0.2	0.2
Marital status								
Married	2619	1191	68.5	65.7	35.0	28.1*	4.8	2.3*
Separated	367	84	74.8	69.2	38.4	19.9	4.3	.8
Divorced	554	221	71.9	69.0	38.5	36.1	7.2	7.9
Widowed	578	220	51.5	41.6	22.3	17.0	2.3	1.9
Never married	1101	341	78.8	69.6*	43.2	35.3	12.1	8.7
Ethnicity								
Black	1947	261	61.6	61.6	29.5	25.8	4.1	3.5
White	1777	1570	71.0	65.9*	37.3	30.2*	6.4	3.5*
Hispanic	1453	150	65.4	66.6	29.5	26.5	4.6	8.9
Other	44	77	64.1	57.0	31.0	21.6	11.1	1.4
Household income								
Above median	1463	913	78.2	73.8	42.6	31.7*	5.8	2.5*
Below median	3405	1031	62.0	56.1*	30.0	25.5	6.7	4.9
Employment status								
Work full-time	2456	1040	78.8	73.7*	44.1	35.2*	8.4	4.3*
Work part-time	452	241	72.6	65.6	34.6	18.9*	5.4	3.6
Retired	717	343	53.0	51.7	23.9	25.1	1.2	2.0
Homemaker	892	227	52.3	44.2	19.8	12.8	1.9	1.1
Other	702	207	66.2	59.9	34.5	30.4	8.5	7.3
Religion								
Catholic	1730	510	82.5	78.6	43.0	37.3	8.5	6.7
Jewish	58	36	88.2	91.8	50.6	30.2	.3	0
Liberal Protestant[b]	412	360	83.4	72.6*	47.6	36.1*	5.4	1.0*
Conserv Protestant[c]	2568	887	53.6	51.1	23.6	19.3	3.6	2.2
Other	437	197	79.3	75.4	48.8	37.1	13.4	9.3
Importance of religion								
Very	3406	1087	56.5	51.5	23.8	18.8*	2.5	1.4
Somewhat	1357	696	84.0	81.0	48.3	39.0*	9.7	5.4*
Not really	283	180	85.8	75.7	53.8	41.6	10.2	9.9
Not at all	112	89	81.9	77.6	54.4	47.1	15.3	9.0
Education								
< High school	2063	479	53.6	50.4	27.8	23.5	5.2	6.3
High school	1628	779	71.9	66.3	35.3	26.4*	9.2	3.8*
Some college	935	401	73.2	70.2	37.3	30.6	4.4	3.1
College graduate	583	397	84.0	75.4*	47.9	39.8	3.4	1.8

(Continued)

report slightly higher rates of alcohol use than did Blacks or Hispanics, but the differences are not significant.

Marital status is a life-style indicator that, together with age, is related to patterns of socializing and to alcohol use. The never married, as a group, are younger than the married, and past surveys have often found that they report higher rates of alcohol use. Similarly, the 1990 data show significant differences between married and never-married respondents in weekly drinking and in the proportion reporting five or more drinks on one occasion at least weekly in the past 12 months.

Measures of household income, education, and employment status are positively associated with most measures of alcohol consumption, although small n's result in nonsignificant differences in many instances. The relationship of household income and the measures of alcohol use is the clearest of these three variables; those reporting household incomes above the median were significantly higher on all six alcohol use measures.

We also contrasted respondents in nonmetropolitan areas with those in metropolitan areas of less than 50 000 population and 50 000 or more. Respondents in large and small metropolitan areas did not differ significantly from each other, but there were significant differences ($P < .05$) between the metropolitan and nonmetropolitan groups on all measures except the five drinks per occasion measure.

Finally, we compared five census divisions of the United States. The Northeast was the "wettest" of these. Compared with the South, the Northeast had significantly higher rates on all consumption measures except the five drinks per occasion measure. The Northeast region's rates were higher than those of the remaining three regions as well; however, few of these differences were statistically significant.

Trends, 1984 through 1990

Table 1 shows data from both 1984 and 1990. In 1990 compared with 1984, there was a significantly lower proportion of current drinkers and a lower proportion of current drinkers of wine, beer, and liquor. There were also significantly fewer respondents reporting that they had drunk any alcoholic beverage as often as once a week in the preceding year, and relatively fewer who reported drinking five or more drinks per occasion as often as once a week.

TABLE 2—Continued

	1984 (n = 5221)	1990 (n = 2058)	Current Drinkers, % 1984	Current Drinkers, % 1990	Weekly Drinkers, % 1984	Weekly Drinkers, % 1990	Drinkers of Five or More Drinks,[a] % 1984	Drinkers of Five or More Drinks,[a] % 1990
Urbanicity								
Metro > 50 000 population	2796	889	72.5	67.3	39.7	31.2*	7.2	4.8
Metro ≤ 50 000 population	850	555	78.6	70.0*	39.6	30.6*	7.1	2.7*
Non-metro	1574	614	63.5	56.3*	31.6	23.7*	4.9	3.7
Region[d]								
Northeast	990	346	80.1	75.5	41.1	37.6	3.4	3.9
Midwest	883	612	74.9	69.3	41.4	27.1*	7.6	3.3*
Pacific	815	266	74.6	64.8	41.7	28.0*	8.5	5.0
South	2322	698	57.8	58.8	26.8	27.1	4.9	3.7
Mountain	211	136	62.1	53.7	31.5	23.2	12.2	4.2

Note. Percentages are based on weighted sample sizes (effective n's of 1228 for the 1984 survey and 1150 for the 1990 survey). Subgroup n's may not equal total because of missing data.

[a]Persons who reported having five or more drinks on one occasion at least once a week during the previous year.

[b]Defined as Protestant (no denomination mentioned), Lutheran, Presbyterian, Episcopalian, Unitarian/Universalist, Quaker, or Congregational.

[c]Defined as Baptist, Methodist, fundamentalist Protestant, Pentecostal, Assembly of God, Church of God, Nazarene, Holiness, Apostolic, Evangelical, Sanctified, Christian Church/Disciples of Christ, United Church of Christ, Christian Reformed, Jehovah's Witness, Seventh Day Adventist, Mormon/Latter Day Saints, Brethren, Spiritualist, Rastafarian, and Salvation Army.

[d]Northeast includes New England and Mid-Atlantic states; Midwest includes East North Central and West North Central states; Pacific includes only Pacific states; South includes South Atlantic, East South Central, and West South Central states; Mountain includes only Mountain states.

*$P < .05$.

Table 2 contains these same measures of alcohol consumption but shown within various categories of demographic characteristics. The same 1984 vs 1990 comparisons were made for the current use of wine, beer, and liquor considered separately (data not shown). In all, we made 336 comparisons, of which 124 were significant ($P < .05$); all these differences were in the direction of less alcohol use in 1990 than in 1984.

Table 2 contains proportions of current drinkers, weekly drinkers (of any alcoholic beverage in the past 12 months), and drinkers of 5 or more drinks on one occasion at least weekly during the past year. In almost all cases, the figures for 1990 are lower than those for 1984, although many differences fail to reach significance. Not shown here are data on current use of the three beverage types asked about in the two surveys. Generally speaking, the same patterns obtain for wine, for beer, and for liquor as obtain for any alcoholic drink. That is, there is a tendency for the 1990 figures to be lower than those for 1984, although the differences in smaller demographic categories are often too small to be statistically significant. As seen in Table 1, the percentage who reported drinking at least five alcoholic drinks at least as often as once a week in the last year was significantly lower in 1990 than in 1984. However, the data shown in Table 2 for persons who drank five or more drinks per occasion suggest that the figures for 1990 are often not significantly lower than those for 1984, even though the direction of differences is almost uniform.

The downturn in alcohol consumption between 1984 and 1990 was fairly general, but not uniformly so. White respondents reported significantly lower alcohol use by all measures in 1990 than in 1984, but no significant differences were found for Black or Hispanic respondents on any of the measures of alcohol use.

Gender differences in alcohol use, although still substantial, were smaller in 1990 than in 1984. For men, the 1990 figures on all six measures of alcohol use were significantly lower than those for 1984; for women, significantly lower figures in 1990 compared with 1984 were found only for alcohol use at least as often as once a week and any current use of beer or liquor.

As in previous surveys, older people reported lower rates of alcohol use than did younger people. Within age categories, the 1990 figures on alcohol use are generally lower than 1984 figures, but the differences often fail to reach statistical significance. The few significant differences between the two surveys occurred more often in the younger age categories (ages younger than 40 years).

Similarly, comparisons of marital status categories show lower rates of alcohol use in 1990 than in 1984, but few differences reach statistical significance. The never married reported more alcohol use than did members of other categories, but it was also among the never married that the rates of alcohol use on all six measures were lower in 1990 than in 1984.

As discussed in the description of the 1990 sample, conservative Protestants reported lower rates of alcohol use on almost all measures than did Jewish, Catholic, and liberal Protestant respondents. Among liberal Protestants, the 1990 rates were significantly lower than those for 1984 on all six measures of alcohol use. Catholics reported significantly lower rates of wine and liquor use in 1990 than in 1984 (data not shown). Respondents who said that religion was "very important" in their lives reported lower rates of alcohol use in 1990 than in 1984.

In general, lower rates of alcohol use were reported across all categories of education, family income, employment status, and urbanicity in 1990 than in 1984, but relatively few differences are statistically significant. One exception is in the rates for those employed full-time, for whom each measure of alcohol use was significantly lower in 1990 than in 1984.

A comparison of 1984 and 1990 data for regions indicates that rates of alcohol use were lower in 1990 than in 1984 except in the South. Of the other four census regions, only the Midwestern region was statistically significantly lower on all measures in 1990 than in 1984.

Given these shifts, is the downturn in alcohol consumption still apparent when the effects of changes in population characteristics are taken into account? To address this question, we used multivariate analyses that included simultaneous entry of all the independent variables shown in Table 2 for each alcohol measure.

The findings from the logistic regressions suggest that the apparent decline in alcohol use among US adults still holds when demographic characteristics of the

population are controlled. The data (not shown) indicate that for each of the six variables predicted, the year of the survey is a highly significant predictor, and that reduced odds of being in each predicted drinking category are associated with 1990 as compared with 1984. For instance, the odds of being a current drinker in 1990 were .696 compared with 1984. Similarly, the odds of being a weekly drinker were .617; of being a current wine drinker, .625; of being a current beer drinker, .703; and of being a liquor drinker, .613. Finally, the odds of reporting drinking five or more drinks on one occasion at least once a week were .515 in 1990 compared with 1984.

Because ethnicity has been an important factor in alcohol use in the United States, separate logistic regressions were run for White, Black, and Hispanic respondents. For Whites, year of survey was significant for all drinking measures except drinking five or more drinks on one occasion at least weekly. Year of survey was not significant for the Black or Hispanic groups for any of the alcohol use measures.

Discussion

The survey data comparing 1990 and 1984 indicate that rates of abstinence have increased in many sociodemographic categories and that some shifts in drinking patterns toward lighter drinking have also taken place. Rates of abstinence among US adults had remained more or less constant for about 50 years,[9] and yet the proportion of respondents who reported abstaining from each beverage or from all beverages seems to have increased in the short time between 1984 and 1990.

However, these data do not suggest that dramatic changes have taken place in the *correlates* of alcohol use since 1984. For instance, men and younger people of both sexes still include more drinkers and heavier drinkers than do women and older people, and regional differences in consumption patterns remain. The downturn in alcohol consumption that occurred among Whites did not extend to Black and Hispanic respondents. Since consumption as measured in 1984 was higher for Whites than for Black and Hispanic respondents, the result is a lessening of the differences between groups, and in 1990 no significant differences remained. Note that in the 1984 but not the 1990 national survey, Blacks and Hispanics were oversampled to permit detailed analyses of drinking patterns and problems in these groups.[10,11]

Will this downward trend in consumption continue? Some researchers have found reason to believe this is the case.[5,12] Room[12] has suggested that the climate of opinion that favored alcohol consumption, and which tolerated some level of alcohol-related problems, reached a periodic high point around 1980 or so and has begun a decline. He notes that a similar downward swing occurred in the years preceding Prohibition, and that this downturn was followed by an upward trend from 1980 to 1993. He argues that it takes about three generations to forget the troubles that accompany high consumption rates, after which consumption may trend upward. And it may take about the same time for rates of alcohol use and problems to increase to the point where popular sentiment supports measures intended to reduce consumption.

Williams et al.[5] add that the population as a whole is aging, and that age is negatively related to alcohol use. Data from the 1990 National Household Survey on Drug Abuse suggest that use of other drugs, including marijuana and cocaine, has declined at about the same time that alcohol use has declined.[13] Given that this trend is occurring, an important issue to consider is whether these changes will affect the prevalence of alcohol-related problems, and, if so, which ones. □

Acknowledgments

This research was supported by a National Alcohol Research Center Grant (AA-05595) from the US National Institute on Alcohol Abuse and Alcoholism to the Alcohol Research Group, California Pacific Medical Center Research Institute, San Francisco, Calif.

This paper was presented at the 120th Annual Meeting of the American Public Health Association, Washington, DC, November 8–12, 1992.

The authors wish to thank Gloria Bocian for her skilled computer programming assistance.

References

1. Clark WB, Hilton ME, eds. *Alcohol in America: Drinking Practices and Problems in a National Survey*. Albany, NY: State University of New York Press; 1991.
2. Johnson P, Armor D, Polich M, Stambul H. *U.S. Adult Drinking Practices: Time Trends, Social Correlates and Sex Roles*. Santa Monica, Calif: RAND Corp; 1977.
3. Clark WB, Midanik L. Alcohol use and alcohol problems among U.S. adults: results of the 1979 survey. In: *Alcohol Consumption and Related Problems*. Washington, DC: US Dept of Health and Human Services; 1982:3–52. DHHS publication ADM 82-1190. Alcohol and Health Monograph 1.
4. Hilton ME, Clark WB. Changes in American drinking patterns and problems, 1967–1984. In: Clark WB, Hilton ME, eds. *Alcohol in America: Drinking Practices and Problems in a National Survey*. Albany, NY: State University of New York Press; 1991: 105–120.
5. Williams GD, Debakey SF. Changes in levels of alcohol consumption, 1983–1988. *Br J Addict*. 1992;87:643–648.
6. Williams GD, Stinson FS, Clem D, Noble J. *Apparent per Capita Consumption: National, State and Regional Trends: 1977–1990*. Washington, DC: US Dept of Health and Human Services; 1992:1–17. DHHS publication ADM 281-89-0001. Surveillance report No. 23.
7. Greenfield TK, Hudes ES, Krotke KP. Practical sampling design issues in national alcohol surveys. Presented at the Alcohol Epidemiology Symposium, Kettil Bruun Society For Social and Epidemiological Research on Alcohol; June 1–5, 1992; Toronto, Canada.
8. Room R. Measuring alcohol consumption in the U.S.: methods and rationales. In: Clark WB, Hilton ME, eds. *Alcohol in America: Drinking Practices and Problems in a National Survey*. Albany, NY: State University of New York Press; 1991:26–50.
9. Hilton ME. Abstention in the general population of the U.S.A. *Br J Addict*. 1986;81:95–112.
10. Caetano R. Findings from the 1984 national survey of alcohol use among U.S. Hispanics. In: Clark WB, Hilton ME, eds. *Alcohol in America: Drinking Practices and Problems in a National Survey*. Albany, NY: State University of New York Press; 1991: 293–307.
11. Herd D. Drinking patterns in the Black population. In: Clark MB, Hilton ME, eds. *Alcohol in America: Drinking Practices and Problems in a National Survey*. Albany, NY: State University of New York Press; 1991: 105–120.
12. Room R. Cultural changes and trends in alcohol problems indicators: recent U.S. experience. In: Clark WB, Hilton ME, eds. *Alcohol in America: Drinking Practices and Problems in a National Survey*. Albany, NY: State University of New York Press; 1991: 149–164.
13. Harrison LD. Measuring drug use in the U.S.A.: overview of the National Household Survey on Drug Abuse. Presented at the 1991 conference of the American Association for Public Opinion Research; May 16–19, 1991; Phoenix, Ariz.

Drinking Water Mutagenicity and Gastrointestinal and Urinary Tract Cancers: An Ecological Study in Finland

Meri Koivusalo, MD, Jouni J. K. Jaakkola, MD, DSc, Terttu Vartiainen, PhD, Timo Hakulinen, ScD, Sakari Karjalainen, MD, Eero Pukkala, MA, and Jouko Tuomisto, MD, PhD

ABSTRACT

Objectives. **The purpose of this study was to investigate the relationship between exposure to mutagenic drinking water and cancers of the gastrointestinal and urinary tract.**

Methods. **Past exposure to drinking water mutagenicity was assessed in 56 Finnish municipalities for the years 1955 and 1970. The cases of bladder, kidney, stomach, colon, and rectum cancers were derived from two periods (1967 to 1976 and 1977 to 1986). Age, sex, social class, urban living, and time period were taken into account in the Poisson regression analysis.**

Results. **Statistically significant exposure-response association was observed between exposure and incidence of bladder, kidney, and stomach cancers. In an ordinary municipality using chlorinated surface water, this exposure would indicate a relative risk of 1.2 for bladder cancer and of 1.2 to 1.4 for kidney cancer compared with municipalities where nonmutagenic drinking water was consumed.**

Conclusions. **The acidic mutagenic compounds present in drinking water may play a role in the etiology of kidney and bladder cancers, but, because the results are based on aggregate data, they should be interpreted with caution. (*Am J Public Health.* 1994;84:1223–1228)**

Introduction

The carcinogenic and mutagenic compounds found in chlorinated drinking water have raised concern over the potential long-term health effects of water chlorination and chlorination by-products.[1–3] Epidemiological studies have suggested an association between chlorinated drinking water and cancers of the urinary and gastrointestinal tract, but the evidence of such a causal relationship has remained inconclusive.[4–6]

The production of chlorination by-products depends on raw water quality and chlorination practices. In epidemiological studies on water chlorination and cancer, the quantitative estimates of exposure are mainly based on the volatile fraction of organic material, which contains compounds such as trihalomethanes. Since most of the organic compounds in drinking water—including the mutagenic ones—are known to be nonvolatile[3] and also acidic,[7] the amount of trihalomethanes may not necessarily reflect the potential carcinogenicity of chlorination by-products found in drinking water.

So far, no epidemiological studies have been published on the association of exposure to mutagenic drinking water and the risk of cancer. Because of the high content of organic material—mainly humic substances—in raw waters, high levels of mutagenic activity have been observed in Finnish chlorinated drinking waters.[7] A major concern is 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone (MX), which has been shown to be a very potent bacterial mutagen[8] and genotoxic compound in laboratory animals.[9] The concentration of MX correlates well with the degree of mutagenicity.[10] MX has also been found in drinking waters in the United States,[11] Great Britain,[12] and Japan.[13] Thus, the purpose of this study was to investigate the relationship between exposure to mutagenic drinking water and risk of cancers of the gastrointestinal and urinary tract.

Subjects and Methods

It was important that the exposed and unexposed persons in the study be as comparable as possible with respect to other cancer risk factors. Thus, municipalities with large rural populations were excluded by limiting the study to those 63 Finnish municipalities administratively defined as cities or rural towns in 1950. To improve the validity of the study, municipalities with fewer than 2000 inhabitants, those with more than 80% of the population born outside the municipality, and those that had been merged with large rural municipalities were excluded from the study, thus reducing the number of municipalities to 56.

Numbers of new cancer cases were obtained from the population-based, na-

Meri Koivusalo, Timo Hakulinen, Sakari Karjalainen, and Eero Pukkala are with the Finnish Cancer Registry, Helsinki. Dr Koivusalo, Terttu Vartiainen, and Jouko Tuomisto are with the Division of Environmental Health, National Public Health Institute. Dr Hakulinen is also with the Unit of Cancer Epidemiology, Karolinska Institute. Jouni J. K. Jaakkola is with the Department of Public Health, University of Helsinki.

Requests for reprints should be sent to Meri Koivusalo, MD, Finnish Cancer Registry, Liisankatu 21 B, 00170 Helsinki, Finland.

This paper was accepted February 1, 1994.

Editor's Note. See related editorial by Cantor (p 1211) in this issue.

TABLE 1—Number of Observed Cancer Cases and Age-Adjusted[15] (World Standard Population) Cancer Incidence per 1 Million Person-Years in the 56 Municipalities, by Sex and Time Period

	No. of Observed Cancer Cases				Cancer Incidence			
	Men		Women		Men		Women	
	1967–76	1977–86	1967–76	1977–86	1967–76	1977–86	1967–76	1977–86
Bladder	1147	1800	467	730	121	145	27	31
Kidney	883	1312	783	1167	91	110	49	56
Stomach	3110	2858	2877	2881	327	233	164	125
Colon	1058	1708	1760	2707	112	139	104	120
Rectum	939	1412	1251	1651	100	114	72	74

TABLE 2—Population Included in the Study, Percentage of Population within the Water Supply System, Average Drinking Water Mutagenicity,[a] Population Exposure,[b] and Range in Mutagenicity[a] in 1955, 1970, and 1985

	1955	1970	1985
Population, in millions	1.7	2.2	2.3
Population within water supply system, %	43	80	92
Average mutagenicity			
All municipalities	1300	1900	1300
All municipalities[c]	2200	2900	1200
Population exposure			
All municipalities	600	1500	1200
All municipalities[c]	1600	2600	1100
Range in mutagenicity	0–5700	0–6900	0–5400

[a]Net revertants per liter (net rev/L).
[b]Percentage of population within the water supply system multiplied by the mutagenicity level.
[c]Weighted by person-years from 1967 to 1986.

tionwide Finnish Cancer Registry. The multiple sources of information available guarantee an almost 100% coverage of all cancer cases in Finland.[14] The study focused on cancers of the stomach, colon, rectum, kidney, and bladder (*International Classification of Diseases*, 7th edition codes 151, 153, 154, 180, and 181, respectively). The observed number of cancer cases was obtained for each municipality by sex, age (5-year categories), and time period (1967 to 1976, 1977 to 1986). The expected number of cancer cases (5-year categories) was calculated for each stratum of sex, age, and period using the respective incidence rates for the total population of Finland. For the analysis, age was further pooled into four broad categories: 0 to 49, 50 to 64, 65 to 74, and from 75 years onward. In addition, age-adjusted incidence rates for all municipalities combined were calculated for both time periods (Table 1).

Assessment of Past Exposure to Mutagenic Drinking Water

The past exposure to drinking water mutagenicity was assessed with a method developed by Vartiainen et al.[16] for two time periods, using the index years 1955 and 1970 to adjust for time and exposure factors such as latency. Information on past routinely collected water quality parameters (e.g., permanganate consumption, pH, color, and ammonium and iron content), water treatment practices, and chemicals used (e.g., chlorine dose, pre- or postchlorination or both) in the municipal waterworks was collected from a questionnaire sent to the municipalities and from past surveys and archives of administrative authorities. The past drinking water mutagenicity was then assessed for both years on the basis of an equation[16] giving the estimated drinking water mutagenicity level in net revertants per liter (net rev/L). The drinking water mutagenicities for the equation were measured according to standard procedures[17] in the Ames test using *Salmonella typhimurium* bacterial tester strains TA100 and TA98 without enzymatic activation.

Water from different waterworks often gets mixed when used by consumers. In those municipalities with more than one waterworks, a mean was calculated weighted for the volumes of water supplied by the different waterworks. The exposure to drinking water mutagenicity was estimated to be zero for people living outside the public water supply system (private wells) as no mutagenicity has been observed in groundwaters.[16] For both years, 1955 and 1970, the population exposure (wm) was calculated by multiplying the proportion of the population within the municipal water supply system (w) by the estimated level of drinking water mutagenicity (m) (Table 2, Figure 1). For the statistical analysis, the level of mutagenicity was also categorized (no mutagenicity, less than 3000 net rev/L, and 3000+ net rev/L), and the average of mutagenicities in 1955 and 1970 was used. A more detailed evaluation and description of the past exposure assessment in terms of drinking water mutagenicity has been published elsewhere.[18]

Statistical Analysis

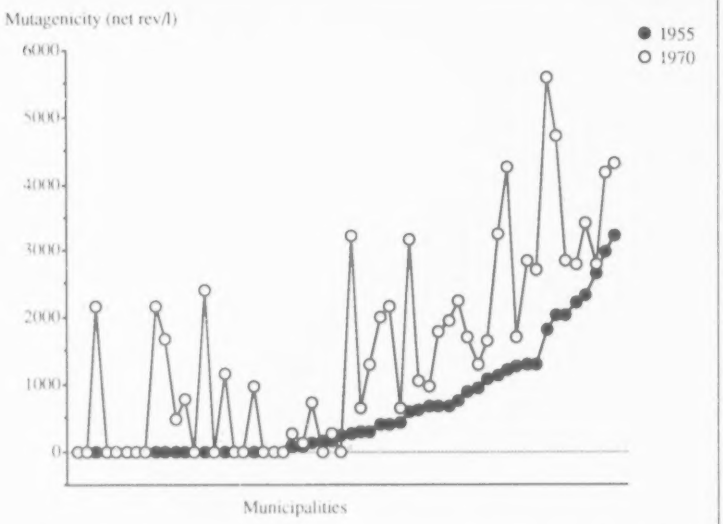
The observed numbers of cancer cases in each municipality were compared with the expected numbers of cancer cases based on the cancer incidence in the entire country in strata defined by sex, broad age group, and calendar period. The risk ratio (RR) between the observed and the corresponding expected number of cancer cases was modeled according to an additive relationship:

$$RR = \alpha + \gamma wm + \Sigma\theta_i s_i.$$

The baseline risk α was allowed to differ in strata defined by sex, age, calendar

period, and urban living. The parameter γ thus gives the change in risk ratio related to a unit of change in population exposure to mutagenicity (wm). In the control of confounding, the relative sizes of social classes in each municipality were used to describe life-style and smoking habits. The relative size of the two highest social classes out of four and the proportion of the lowest class in each municipality were derived from the 1970 census[19] and were used in the analysis as numerical variables (s_i, i = 1,2), with parameters θ_i specifying the relationship with social classes. The three largest cities in Finland, with their comparably more urban environments and life-styles, were grouped into a separate category in the analysis, and a categorical variable was created to adjust for a different base level of risk in these cities.

The statistical analysis was performed using GLIM[20] with an identity link function and a statistical model based on Poisson regression.[21] The effect of drinking water mutagenicity was studied as both a categorical and a numerical variable, and logarithmic transformations were also fitted. In the final analysis, the choice of the model with respect to mutagenicity was based on its ability to explain the variation in the data with fewest model parameters. This resulted in a linear model with a numerical variable for exposure to mutagenicity in the final analysis. Overdispersion was estimated to be of the order of 10%.[22] Adjusting for this would not have changed the confidence intervals (CIs) of the model parameters to any substantial extent. For purposes of illustration, the transformed coefficient, $1 + \gamma 3000$ net rev/L, was used to derive a risk ratio related to a typical mutagenicity exposure of 3000 net rev/L (Table 2). In further study of confounding, separate analyses were performed without the three largest cities, without eight municipalities that had 2% or more of the population working in the chemical industry,[23] and without six municipalities that had the highest levels of migration (an average annual in-migration of more than 8% from 1966 to 1970).[24] To focus the analysis on the exposure-response association and the quantity of the biologically active compounds in drinking water, the statistical analysis was also performed separately for those 34 municipalities in which the population was exposed to mutagenic drinking water. The proportions of cancer cases attributable to exposure were calculated[25] using the linear model and the average population exposure level of 1600 net rev/L in 1955 (Table 2).



Note. Exposures were estimated by multiplying the proportion of the population within the municipal water supply system by the mutagenicity level.

FIGURE 1—Estimated exposures to drinking water mutagenicity in the 56 municipalities in 1955 and 1970, arranged by the estimated exposures in 1955.

Results

Considering the population mutagenicity exposure as the only risk factor in the model, the risk ratios of cancers of the bladder, kidney, and colon were significantly higher than 1 (Table 3). In an average municipality with chlorinated surface water, the exposure to drinking water mutagenicity for those within the municipal water supply system was 2700 net rev/L in 1955 and 3300 net rev/L in 1970. Thus, in colon cancer, a risk ratio of 1.31 corresponded to an exposure of 3000 net rev/L, indicating a 10% unit increase for each addition of 1000 net rev/L, according to the linear exposure-response relationship used. After adjusting for the background variables (age, sex, time period, largest cities) and social class, the risk ratios diminished in kidney and colon cancers.

The risk of bladder cancer was 1.17-fold (95% CI = 1.03, 1.31) among exposed as compared with unexposed persons when adjusted for the potential confounders. The risk remained approximately the same when the analysis was restricted to the 34 municipalities with mutagenic drinking water. The adjusted risk ratio for kidney cancer was 1.08 (95% CI = 0.91, 1.24) in the total study population and 1.32 (95% CI = 1.11, 1.53) in the restricted analysis. The risk of stomach cancer was also slightly higher among the exposed, with an adjusted risk ratio of about 1.1. The adjusted risk of colon and rectum cancer among exposed persons did not differ from the corresponding risk for unexposed persons by more than could be expected by chance.

The exposure was assessed for 1955, and the cancer risk was calculated for 1967 to 1986 (Table 3). Taking into account cases diagnosed between 1977 and 1986 only, the adjusted risk ratios were approximately the same, with the exception of slightly higher risks in kidney cancer (RR = 1.16; 95% CI = 0.97, 1.36) (Table 4). Applying the average exposure from 1955 and 1970 and cases diagnosed between 1977 and 1986, the adjusted risk ratios for bladder cancer were lower than those given in Table 3 (RR = 1.05; 95% CI = 0.92, 1.19 for all municipalities). In contrast to bladder cancer, slightly higher risks were observed in stomach cancer (RR = 1.09; 95% CI = 1.02, 1.17 for all municipalities). The restricted analysis of

TABLE 3—Estimated Risk Ratio for a Typical Exposure to Chlorinated Surface Water (3000 net rev/L)

Site and Model[b]	All[a] Risk Ratio	All[a] 95% Confidence Interval	Mutagenic Water[a] Risk Ratio	Mutagenic Water[a] 95% Confidence Interval
Bladder cancer				
Mutagenicity	1.18	1.10, 1.27	1.21	1.10, 1.32
Plus age, sex, period, main cities	1.20	1.08, 1.33	1.24	1.09, 1.38
Plus social class	1.17	1.03, 1.31	1.21	1.02, 1.40
Kidney cancer				
Mutagenicity	1.37	1.27, 1.47	1.57	1.44, 1.69
Plus age, sex, period, main cities	1.23	1.10, 1.37	1.44	1.29, 1.60
Plus social class	1.08	0.91, 1.24	1.32	1.11, 1.53
Stomach cancer				
Mutagenicity	0.97	0.93, 1.01	0.99	0.93, 1.04
Plus age, sex, period, main cities	1.03	0.97, 1.09	1.06	0.99, 1.14
Plus social class	1.07	1.00, 1.13	1.11	1.02, 1.19
Colon cancer				
Mutagenicity	1.31	1.23, 1.39	1.39	1.30, 1.48
Plus age, sex, period, main cities	1.18	1.07, 1.27	1.25	1.13, 1.38
Plus social class	1.05	0.93, 1.17	1.09	0.93, 1.25
Rectal cancer				
Mutagenicity	1.05	0.97, 1.12	1.02	0.93, 1.12
Plus age, sex, period, main cities	1.02	0.92, 1.12	0.94	0.87, 1.12
Plus social class	1.05	0.93, 1.17	0.98	0.82, 1.14

[a]The risk ratios for cancer cases in 1967–1986 are calculated corresponding to exposure to drinking water mutagenicity in 1955 for all municipalities (n = 56) and municipalities with mutagenic drinking water (n = 34).
[b]Within each type of cancer, the variables listed are those added to the previous model.

municipalities with mutagenic drinking water resulted in slightly higher risk ratios for bladder, kidney, and stomach cancers. For other cancers, the change of time and exposure periods led to no major differences from the risk ratios reported above.

The analyses performed after exclusion of the three largest cities and the municipalities with the highest migration levels did not change the observed risk estimates. In kidney cancer, slightly higher risk ratios were observed after exclusion of the municipalities with chemical industry (Table 4). Interaction between exposure and gender was statistically significant ($P < .05$) only in kidney and colon cancers. For colon cancer, the adjusted risk ratios corresponding to those in Table 3 were 1.14 (95% CI = 0.92, 1.36) for men and 1.01 (95% CI = 0.87, 1.16) for women. The interaction between age and exposure was statistically significant ($P < .05$) in kidney cancer, with the greatest risk ratio observed among the oldest age group (Table 4).

The exposure-response association may be converted into the proportion of cancer cases from 1967 to 1986 attributable to drinking water mutagenicity. When the adjusted relative risks from the restricted analysis of municipalities with mutagenic drinking water in Table 3 are used, the calculated attributable proportions would be 10% in bladder, 15% in kidney, and 5% in stomach cancer. The respective numbers of new cancer cases between 1967 and 1986 may be derived by applying these proportions to the numbers of cases reported in Table 1.

Discussion

The results indicate that, in a typical municipality using chlorinated surface water (3000 net rev/L), the relative risks would be about 1.2 for bladder cancer, 1.2 to 1.4 for kidney cancer, and 1.1 to 1.2 for stomach cancer as compared with those persons not using mutagenic drinking water. When the exposed population is large, the detection and estimation of even small risks is important from the public health point of view. In the present study, the existence of a method for assessing the past exposure to drinking water mutagenicity, an accurate system of cancer registration, and a large variation in drinking water mutagenicities made it possible to demonstrate small, yet statistically significant excess cancer risks associated with the assessed past exposure to mutagenic drinking water.

The observation of an increased risk of bladder cancer is consistent with results of earlier case-control studies on drinking water chlorination and cancer.[4–6,26] An association between the exposure to chlorinated water and the occurrence of gastrointestinal cancers has also been reported in epidemiological studies.[4–6,27] Such an association was not observed in rectum and colon cancers in our study. In kidney and colon cancers, the observed difference in risk estimates by sex could be explained by higher consumption of water by men. Confounding by gender-related life-styles and occupations or interaction with other gender-related exposures and biological factors may also be involved. The observation of substantially higher relative risks for kidney cancer in the oldest age group is in line with a longer latency time. The observed increased risks of kidney and bladder cancers are in line with the pharmacokinetics of the strong water mutagen MX.[28] The protein-bound MX[29] may be released in the kidney and therefore may induce higher concentrations there than in most other tissues.[28] Although, according to current knowledge, MX is the compound of major concern, chlorinated compounds other than MX may also have a carcinogenic effect.

Socioeconomic status has been shown to be associated with the incidence of several cancers in Finland, including cancers of the colon, rectum, and kidney.[30,31] In the present study, social class was determined by occupation. In earlier studies, it has been shown to be associated with diet and way of life.[32] Occupation has been shown to be associated with smoking, explaining a large part of the variation in lung cancer incidence.[33] Therefore, social class may be regarded as a suitable surrogate variable for life-style, diet, and smoking on the population level. After the municipalities with the highest proportions of the population working in the chemical industry were excluded, the risk estimates were higher in bladder and kidney cancers. This finding may be owing to the confounding effect of chemical exposures in these municipalities. The population density is greater and the life-style different in the three largest cities. After social class variables were included in the model, the risk estimates were similar in the analyses with or without the three largest cities, which

TABLE 4—Risk Ratios for a Typical 3000 net rev/L Exposure in 1955 for Kidney Cancer When Background Factors and Social Class Are Included in the Model, Shown by Time Period and Chosen Restrictions

	Cancer Cases in 1967–1986				Cancer Cases in 1977–1986			
	All		Mutagenic Water		All		Mutagenic Water	
Restrictions	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
No exclusions	1.08	0.91, 1.24	1.32	1.11, 1.53	1.16	0.97, 1.36	1.44	1.18, 1.69
No chemical industry	1.17	1.01, 1.34	1.37	1.15, 1.59	1.26	1.06, 1.47	1.50	1.24, 1.76
Women	0.91	0.69, 1.13	1.21	0.93, 1.50	0.95	0.68, 1.22	1.28	0.94, 1.62
Men	1.26	1.02, 1.49	1.45	1.14, 1.75	1.39	1.10, 1.67	1.60	1.24, 1.97
Age group 75+	1.61	0.92, 2.30	2.46	1.62, 3.30	1.79	0.98, 2.61	2.52	1.50, 3.53

Note. RR = risk ratio, CI = confidence interval.

indicates that the confounding effect of urban living was taken into account. As the inclusion of social class in the multiple regression model may not fully adjust for life-style factors, the residual confounding from life-style factors such as smoking may not be totally excluded.

In a previous study of 86 drinking water samples, the correlation coefficient between the measured mutagenicity in TA100 Ames *S. typhimurium* tester strain and the modeled mutagenicity was 0.85.[16] Assessment of past exposure is based on historical observed raw water quality parameters and water treatment practices and therefore on the real historical situation, even though the actual historical mutagenicity may only be approximated. The methods for determining permanganate values of water samples have remained unchanged since the 1930s. In surface waterworks, the water quality data have been analyzed monthly; in ground waterworks they are analyzed usually once every four seasons. The approximation of historical mutagenicity was based on annual averages, derived from monthly measurements in surface waterworks. These values therefore represent potential seasonal variations.

The current understanding is that the mutagenic compounds in drinking waters are predominantly produced during water chlorination.[34] In an earlier survey, only one raw water source was shown to be mutagenic in the 1980s, and even in this case the mutagenicity of drinking water grew several times higher during water treatment.[16] Other exposures mediated by the water supply network, such as radioactive substances or asbestos from raw water or waterpipes, are unlikely,[35,36] and there is no evidence that these are related to drinking water mutagenicity. Until recently, consumption of bottled water and soft drinks has been very low in Finland, at least when compared with that in western and southern Europe and the United States.

The good coverage and quality of the Finnish Cancer Registry ensure the validity of the outcome assessment. Finland has a rather homogeneous population. During the 1960s, the direction of migration was from rural areas to towns, and therefore from areas of no or low exposure to high-exposure regions. In the 1970s and 1980s, migration between towns was also substantial. Confounding by migration would most probably diminish the estimated risks ratios.

Ecological studies are known to involve problems such as the ecological fallacy.[37] In the present study, the proportion of exposed persons was taken into account in each municipality, thus reducing the degree of ecological fallacy but not eliminating it, as the analysis was based on aggregate data. The fact that the expected rates were related to the whole country rather than to the chosen municipalities was taken into account by including in the model baseline risk parameters specific for sex, age, time period, and urban living. The Poisson distribution assumption of the number of observed cases enabled us to take into account the different sizes of the municipalities, and to study the proportion of persons exposed and the mutagenicity level of those exposed simultaneously in the analyses. The observed exposure-response association in the restricted analysis of municipalities with exposure to mutagenic drinking water indicated that the observed risk is more closely associated with the level of mutagenicity than with drinking water chlorination or mutagenicity as such.

Our findings suggest that the acidic mutagenic compounds present in chlorinated drinking water may play a role in the etiology of kidney, bladder, and possibly stomach cancer. Nevertheless, as the study was based on aggregate data, our results should be interpreted with caution. □

Acknowledgments

This study received financial support from the Academy of Finland, the Yrjö Jahnsson Foundation, the Emil Aaltonen Foundation, and the Finnish Cancer Institute.

References

1. Bellar TA, Lichtenberg JJ, Kroner RC. The occurrence of organohalides in chlorinated drinking waters. *J Am Water Works Assoc.* 1974;66:703–706.
2. Rook JJ. Formation of haloforms during chlorination of natural waters. *Water Treatment Exam.* 1974;23:234–243.
3. Cheh AM, Skochdopole J, Koski P, Cole L. Nonvolatile mutagens in drinking water: production by chlorination and destruction by sulfite. *Science.* 1980;207:90–92.
4. Morris RD, Audet A-M, Angelillo IF, Chalmers TC, Mosteller F. Chlorination, chlorination-byproducts and cancer: a meta-analysis. *Am J Public Health.* 1992;82:955–963.
5. Chlorinated drinking water; chlorination by-products; some other halogenated compounds; cobalt and cobalt compounds. Lyon, France: International Agency for Research on Cancer; 1991. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 52.
6. Murphy A, Craun G. A review of recent epidemiologic studies reporting associations between drinking water disinfection and cancer. In: Jolley R, Condie L, Johnson JD, et al., eds. *Water Chlorination: Chemistry, Environmental Impact and Health Effects.* Chelsea, Mich: Lewis Publishers, Inc; 1990;6:361–372.
7. Vartiainen T, Liimatainen A. High levels of mutagenic activity in chlorinated drinking water in Finland. *Mutat Res.* 1986;169:29–34.
8. Holmbom B, Voss R, Mortimer RD, Wong A. Fractionation, isolation and character-

ization of Ames mutagenic compounds in kraft chlorination effluents. *Environ Sci Technol.* 1984;18:333–337.

9. Jansson K, Mäki-Paakkanen J, Vaittinen S-L, Vartiainen T, Komulainen H, Tuomisto J. Cytogenetic effects of 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone (MX) in rat peripheral lymphocytes in vitro and in vivo. *Mutat Res.* 1993;299:25–28.
10. Kronberg L, Vartiainen T. Ames mutagenicity and concentration of the strong mutagen 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone and its geometric isomer E-2-chloro-3-(dichloromethyl)-4-oxo-butenoic acid in chlorine treated tap water. *Mutat Res.* 1988;206:177–182.
11. Meier JR, Knohl RB, Coleman WE, et al. Studies on the potent bacterial mutagen, 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone: aqueous stability, XAD recovery and analytical determination in drinking water and in chlorinated humic acid solutions. *Mutat Res.* 1987;189:363–373.
12. Horth H, Fielding M, Gibson T, James HA, Ross H. *Technical Report (PRD 2038-M).* Great Britain: Water Research Centre; 1989.
13. Suzuki N, Nakanishi J. The determination of strong mutagen 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone in drinking water in Japan. *Chemosphere.* 1990;21:387–392.
14. Saxén E, Teppo L. *Finnish Cancer Registry 1952–1977.* Helsinki, Finland: Finnish Cancer Registry; 1978.
15. Muir C, Waterhouse J, Mack T, et al., eds. *Cancer Incidence in Five Continents.* Vol 5. Lyon, France: International Agency for Research on Cancer; 1987; 5. IARC Scientific Publications 88.
16. Vartiainen T, Liimatainen A, Kauranen P, Hiisvirta L. Relations between drinking water mutagenicity and water quality parameters. *Chemosphere.* 1988;17:189–202.
17. Maron DM, Ames BN. Revised method for the *Salmonella* mutagenicity test. *Mutat Res.* 1983;113:173–215.
18. Koivusalo M, Vartiainen T, Jaakkola JJK. Drinking water mutagenicity in past exposure assessment of the studies on drinking water and cancer—application and evaluation in Finland. *Environ Res.* 1994:64:90–101.
19. *Population Census 1970:* Vol 9. *Occupation and Social Position;* Vol 16. *Statistical Maps.* Helsinki, Finland: Central Statistical Office of Finland; 1975. 6C104.
20. Payne CD, ed. *The GLIM System Release 3.77. Generalized Linear Interactive Modelling Manual.* Oxford, England: Numerical Algorithms Group; 1986:1–183.
21. Breslow NE, Day NE. *Statistical Methods in Cancer Research:* Vol 2. *The Design and Analysis of Cohort Studies.* Lyon, France: International Agency for Research on Cancer; 1987. IARC Scientific Publications 82.
22. McCullagh P, Nelder JA. *Generalized Linear Models.* 2nd ed. Cambridge, England: University Press; 1989. Monographs on Statistics and Applied Probability 37.
23. *Population Census 1950:* Vol 2. *Population by Industry and Industrial Status.* Helsinki, Finland: Central Statistical Office of Finland; 1955. VI C 102.
24. Herva A. *SEDDATA. Database Consisting of Sociologic, Economic and Demographic Variables.* Helsinki, Finland: University of Helsinki, Department of Statistics; 1977. Research Report 9.
25. Lubin JH, Boice JD. Estimating Rn-induced lung cancer in the United States. *Health Phys.* 1989;57:417–429.
26. Cantor KP, Hoover R, Hartge P. Bladder cancer, drinking water source and tap water consumption: a case-control study. *J Natl Cancer Inst.* 1987;79:1269–1279.
27. Flaten TP. Chlorination of drinking water and cancer incidence in Norway. *Int J Epidemiol.* 1992;21:6–15.
28. Komulainen H, Vaittinen S-L, Vartiainen T, Lötjönen S, Paronen P, Tuomisto J. Pharmacokinetics of 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone (MX), a drinking water mutagen, after a single dose in the rat. *Pharmacol Toxicol.* 1992;70:424–428.
29. Haataja L, Vartiainen T, Lampelo S, Lötjönen S, Tuomisto J. Binding of the strong mutagen, 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone (MX) to bovine serum albumin. *Toxicol Lett.* 1991;59:187–195.
30. Pukkala E, Teppo L. Socioeconomic status and education as risk determinants of gastrointestinal cancer. *Prev Med.* 1986;15:127–138.
31. Rimpelä AH, Pukkala E. Cancers of affluence: positive social class gradient and rising incidence trend in some cancer forms. *Soc Sci Med.* 1987;24:601–606.
32. Aro S, Räsänen L, Telama R. Social class and changes in health-related habits in Finland in 1973–1983. *Scand J Soc Med.* 1986;14:39–47.
33. Pukkala E, Teppo L, Hakulinen T, Rimpelä M. Occupation and smoking as risk determinants of lung cancer. *Int J Epidemiol.* 1983;12:290–296.
34. Wigilius B, Boren H, Grimwall A, Carlberg GE, Hagen I, Brogger A. Impact of bleached kraft mill effluents on drinking water quality. *Sci Total Environ.* 1988;74:75–96.
35. Salonen L, Saxén R. Talousvesien radioaktiivisuus Suomessa. (Radioactivity of Finnish drinking waters.) *Kemia-Kemi.* 1991;5:400–406.
36. Asbestikuidut vedessä. Helsinki, Finland: Suomen Kaupunkiliitto; 1987. Kaupunkiliiton julkaisu 127. (Asbestos in Finnish drinking waters. Publication of the Finnish Association of Cities, no. 127.)
37. Robinson WS. Ecological correlations and the behavior of individuals. *Am Sociol Rev.* 1950;15:351–357.

A National Study of Transitional Hospital Services in Mental Health

Robert A. Dorwart, MD, MPH, and Claudia W. Hoover, MA

ABSTRACT

Objectives. **Shifts in care for the seriously mentally ill from inpatient to community-based treatment have highlighted the importance of transitional care. Our objectives were to document the kinds and quantity of transitional services provided by psychiatric hospitals nationally and to assess the impact of hospital type (psychiatric vs general), ownership (public vs private), case mix, and revenue source on provision of these services.**

Methods. **A national sample of nonfederal inpatient mental health facilities (n = 915) was surveyed in 1988, and data were analyzed by using multiple regression.**

Results. **Half (46%) of the facilities surveyed provided patient follow-up of 1 week or less, and almost all (93%) conducted team review of discharge plans, but 74% provided no case management services. Hospital type was the most consistent predictor of transitional care, with psychiatric hospitals providing more of these services than general hospitals. Severity of illness, level of nonfederal funding, urbanicity, and teaching hospital affiliation were positively associated with provision of case management.**

Conclusions. **Transitional care services for mentally ill patients leaving the hospital were found to be uneven and often inadequate. Reasons for broad variation in services are discussed. (*Am J Public Health.* 1994;84:1229–1234)**

Introduction

One result of the deinstitutionalization of patients from public mental hospitals in the 1970s was the movement of many patients with severe and persistent psychiatric disorders into general hospitals, effectively mainstreaming the mentally ill into the acute medical care system.[1] Another result was the reorientation of the locus of treatment for less acutely ill patients from inpatient facilities to a myriad of community-based clinical and support services.[2] There is general consensus among clinical and policy researchers, as well as among service providers, that although initial efforts to make this transition were partially successful, they were also marked by failures, as evidenced by rising readmission rates of the mentally ill to hospitals and a significant number of severely mentally ill individuals among the growing population who were homeless.[3] These developments brought a recognition of the lack of horizontal and vertical integration of administrative and financial functions within the mental health care system and acknowledgment that availability of services had been mistakenly equated with access to those services. By the 1980s, there was an emphasis on introducing managed care and central mental health authorities, which were expected to improve continuity and appropriateness of treatment.[4,5] In the last decade there has also been a shift from a predominantly public responsibility for inpatient mental health care to an increase in the number of patients cared for in private psychiatric hospitals.[6]

The Problem of Continuity of Care

Changes in the delivery system from public to private providership and from specialty hospitals to general hospitals and community-based services have raised concerns about the accessibility and quality of mental health care, particularly for patients with chronic disorders.[7,8] In response, both policymakers and service providers have targeted the importance of continuity of care for both clinical and supportive services. The transition from inpatient to community-based care is a particularly important one for psychiatric patients because it is at this point in the treatment system that many patients fail to connect with posthospital caregivers and fail to establish meaningful ties to needed aftercare services.

Although providing continuity of care is considered a critical function in the successful treatment of the seriously mentally ill, little is known about the kind and quantity of such services available to psychiatric inpatients in the United States.[9,10] This study attempted to discover the nature and amount of transitional services available to the seriously mentally ill when they leave inpatient care to live in the community. In this article, we distinguish transitional services—those specifically designed to link inpatient and ambulatory care—from traditional aftercare services provided in community-based settings.

The authors are with the Malcolm Wiener Center for Social Policy, John F. Kennedy School of Government, Harvard University, Cambridge, Mass, and Harvard Medical School Department of Psychiatry (The Cambridge Hospital).

Requests for reprints should be sent to Robert A. Dorwart, MD, MPH, Malcolm Wiener Center for Social Policy, 458 Taubman Bldg, John F. Kennedy School of Government, 79 John F. Kennedy St, Cambridge, MA 02138.

This paper was accepted November 3, 1993.

The Link between Continuity of Care and Clinical Outcome

Numerous studies have documented the importance of aftercare services in improving the quality of life and reducing clinical symptoms of former psychiatric inpatients. Despite methodological shortcomings in some studies, the overall weight of evidence suggests that recently discharged inpatients who receive some kind of aftercare (e.g., medications psychotherapeutic, occupational, or case management) will function better after leaving the hospital than those who do not.[11-13]

Although it has been shown that aftercare services have beneficial effects, getting patients to aftercare is a need that has been largely unmet. Myerson and Herman concluded from their review of aftercare studies that "bridging strategies" were the most important link to satisfactory aftercare.[14] Despite its importance, the transition to aftercare is fraught with geographic, interpersonal, temporal, and logistical obstacles.[12,15]

In recognition of these difficulties, model treatment programs, such as the Robert Wood Johnson nine-city Demonstration Program,[4] have been designed specifically to minimize the gaps in treatment and support that characterize the majority of transitions from hospital to community-based services. To do this, model programs provide centrally administered financial, social, and clinical services to the patient. In most local mental health care systems, however, such central administration is not currently in use. In these systems, continuity of care occurs only when a case manager system is in place. Although models of case management vary, the goal in each is to provide continuity of care between treatment settings and to assist in matching services to the needs of each client. Case management is widely regarded as the most effective way to maximize patients' gains.[16-18]

Despite this evidence, such services are scarce. A New York State study of discharge planning in psychiatric facilities found that a large percentage of inpatients were referred to mental health clinics (85%), but very few facilities offered case-management services.[19] According to this survey, public specialty hospitals tend to provide more adequate discharge planning than do general hospitals. Services offered included formal discharge planning, psychiatric appointments, day program referral, and monitoring of patient care for more than 6 months after discharge. Shortell et al.[20] surveyed not-for-profit and for-profit system-affiliated hospitals and found that not-for-profit hospitals offered a higher volume of alternatives to inpatient services, including more unprofitable services, than did for-profit hospitals. Clark and Fox demonstrated the cost-effectiveness of case-management services.[21]

Hypotheses

Based on previous research, we anticipated that two structural measures—hospital type (specialty or general) and ownership (private or public)—would influence provision of transitional services. We expected to find that psychiatric (specialty) hospitals would provide transitional care to a greater extent than general hospitals for several reasons. First, the case mix of specialty hospitals (especially public ones) tends to include more chronic and severely mentally ill patients.[20] These patients are more likely to require transitional care of greater intensity and duration than that required by acutely ill, short-term patients usually seen in the psychiatric wards of general hospitals. Second, administratively, it is often more feasible for specialty hospitals than for general hospital units to provide transitional care services; follow-up and transitional care can be time consuming for staff and are more easily organized when a high proportion of patients with psychiatric difficulties justifies the assignment of full-time staff for this purpose. We also anticipated that public facilities would provide more transitional services than would private ones. Ownership form has been shown to be related to institutional mission, which is, in turn, correlated with institutional behavior.[5] Public facilities are expected to see their mission as one primarily of stewardship rather than profitability, thus allowing administrators to make decisions about service mix and staffing based on patient needs.

We anticipated, too, that source of revenues might be related to the provision of transitional services. It has been noted previously that one way in which privatization has affected mental health care providers is that, regardless of ownership, almost all inpatient facilities receive revenue from a variety of sources: contracts with states and counties, Medicaid, Medicare, and private insurers. We expected that those facilities receiving a greater proportion of public revenues would provide the most transitional care. The diagnostic mix of patients treated would also be related, we thought, to transitional services. Patients with long-term chronic illness, such as schizophrenia, often have greater need for assistance in making the transition to outpatient or community-based facilities than do those with shorter-term difficulty, such as depression.

Methods

The National Mental Health Facilities Study,[22] conducted between October 1987 and 1988, was designed to address unanswered questions about the ongoing privatization of mental health services in the United States. The survey consisted of a 200-item questionnaire mailed to some 900 administrators of all nonfederal psychiatric hospitals in the United States, including community mental health agencies with inpatient units, and a 75% random sample of psychiatric units in nonfederal general hospitals. The overall response rate was 60%, ranging from 78% for public specialty hospitals to 38% for for-profit specialty hospitals. A comparison of our findings on revenue sources, case mix, and staffing practices of for-profit specialty hospitals with data from national surveys by the National Institute of Mental Health[23] and the National Association of Private Psychiatric Hospitals yielded overall consistency. A more detailed discussion of sample selection and response rate was published previously.[22]

Questions on the survey addressed a broad range of issues concerning hospital ownership status, referral sources, clinical services, patient characteristics including diagnostic mix and payer mix, staffing policies, and treatment monitoring policies. Additional information about area characteristics for each hospital, such as population density, was available from the Department of Commerce's Area Resource File.[24] Within this context, the National Mental Health Facilities Study examined the impact of hospital type (specialty vs general), hospital ownership (public vs private), hospital revenue sources, and patient diagnostic mix on the type and duration of transitional care given discharged inpatients.

Dependent Variables

Hospital administrators were asked about six different aspects of their facilities' transitional care practices. A hospital's services included those provided by hospital staff and those contracted through other providers with hospital-administered funds. First, they were asked how

often (on a scale of never, sometimes, often) primary responsibility for ensuring that patients receive aftercare (e.g., psychotherapy, occupational, or case-management services) rested with the hospital as compared with the patients themselves, private practitioners, and state and community agencies; second, whether their facility conducted team review of discharge plans; third, whether staff were exclusively assigned to patient follow-up; fourth, whether case-management services were provided; fifth, whether the average length of follow-up contact lasted 1 week, up to 1 month, up to 6 months, or more than 1 year; and sixth, which of six follow-up methods was the one most often used: (1) patient is given follow-up appointment; (2) patient is given staff help with referral; (3) patient is sent appointment reminder; (4) patient is telephoned by aftercare provider; (5) staff visits aftercare providers; and (6) staff visits patients' homes.

Independent Variables

Each hospital reported the legal form of ownership (public, private nonprofit, private for-profit) under which it operated, as well the type of psychiatric unit (general or specialty) and whether the unit was a teaching facility for medical interns and psychiatric residents.

Hospital administrators reported the percentage of total psychiatric patient care revenues from private insurers, Medicare, Medicaid, state and county contracts, and direct payment by patients. Administrators also reported the proportion, by diagnostic category, of inpatients treated in the last year. The percentage of patients diagnosed with schizophrenia was used as an indicator of long-term chronic illness and thus identified the hospitals with a high proportion of those patients most likely to need specific help in making the transition from inpatient care to community-based care (see Table 1).

Statistical Analysis

Several variables were used initially in regression analyses to control for hospital and community characteristics. Generally, the availability of resources of many kinds (including staff and level of staff training) is greater for larger hospitals and for those in more urban areas. To control for these influences, we included a measure of the percentage of the hospital service area population that is urban (Department of Commerce Area Resource File),[24] the average daily hospital psychiatric census, and the concentration

TABLE 1—Descriptive Statistics for Independent Variables

	No. of Hospitals	% of Sample
Categorical variables		
Type		
Specialty	349	38.1
General	566	61.9
Ownership		
Public	289	31.6
Private	626	68.4
Residency program		
No	718	78.4
Yes	189	20.6
		Mean ± SD
Continuous variables		
Percentage of patients diagnosed with schizophrenia	902	27.37 ± 18.78
Percentage of hospital revenues from Medicaid	888	18.94 ± 18.92
Percentage of hospital revenues from Medicare	888	20.96 ± 14.88
Percentage of hospital revenues from state and county	886	10.42 ± 23.93
Daily hospital psychiatric census	908	103.64 ± 209.60
Population density in hospital service area	915	75.78 ± 21.02

Note. Response rates to individual questions vary somewhat.

TABLE 2—Transitional Care Services Provided by Psychiatric Hospitals and General Hospital Psychiatric Units (n = 915)

	Frequency, %		
	Psychiatric Hospital (n = 349)	General Hospital (n = 566)	Total (n = 915)
Hospital responsible for ensuring aftercare			
Never	12	17	15
Sometimes	31	38	35
Often	57	45	50
Length of follow-up			
Up to 1 week	32	55	46
Up to 1 month	27	27	27
Up to 6 months	18	10	13
Up to 1 year	23	8	14
Method of follow-up used most often			
Patient given appointment	49	66	59
Staff available to help with referral	19	19	19
Letter sent reminding of patient appointment	4	4	4
Phone follow-up with aftercare provider	15	8	11
Staff visit to aftercare provider	7	2	4
Staff visit to patient's residence	6	1	3
Team review of discharge plan			
No	3	10	7
Yes	97	90	93
Staff assigned exclusively to patient follow-up			
No	30	48	41
Yes	70	52	59
Case management provided			
No	65	79	74
Yes	35	21	26

Note. Response rates to individual questions vary somewhat.

TABLE 3—Regression Analysis of Hospital Characteristics Associated with Providing Transitional Services for Psychiatric Patients

	Dependent Variable									
	Degree of Responsibility for Ensuring Aftercare[a]		Length of Patient Follow-Up[b]		Intensity of Patient Follow-Up[c]		Staff Assigned to Follow-Up (0 = No; 1 = Yes)		Case Management Provided (0 = No; 1 = Yes)	
Independent Variable	β	*P*	β	*P*	β	*P*	Odds Ratio	95% CI	Odds Ratio	95% CI
Hospital type (0 = specialty; 1 = general)	−0.223	.0003	−0.705	.0001	−0.693	.0001	0.26*(−)[d]	0.11, 0.60	0.74	−0.71, 1.12
Hospital ownership (0 = public; 1 = private)	−0.072	.2859	−0.082	.3945	−0.032	.7978	0.66	0.31, 1.42	1.17	0.75, 1.84
Percentage of patients diagnosed with schizophrenia	0.001	.6355	−0.005	.0413	−0.005	.0954	1.00	0.98, 1.02	1.02*(+)	1.01, 1.03
Percentage of hospital revenues from Medicaid	−0.002	.2636	−0.002	.4596	−0.001	.7377	1.00	0.99, 1.02	1.00	0.99, 1.01
Percentage of hospital revenues from Medicare	−0.000	.9037	0.001	.6408	0.000	.9257	1.00	0.98, 1.02	1.01	1.00, 1.03
Percentage of hospital revenues from state and county	−0.003	.0299	−0.003	.1699	−0.004	.1535	0.99	0.97, 1.00	1.01*(+)	1.00, 1.02
Residency program (0 = no; 1 = yes)	0.125	.0509	0.182	.0488	0.233	.0664	0.61	0.30, 1.26	0.57*(+)	0.38, 0.86
Daily hospital psychiatric census	−0.000	.0877	−0.000	.3815	0.000	.5568	1.00	1.00, 1.00	1.00	1.00, 1.00
Percentage of urban population in service area	0.003	.0171	0.002	.3538	0.003	.1880	0.99	0.98, 1.00	1.00	0.99, 1.00

Note. Data in ordinal form were treated as continuous and analyzed by using multivariate least squares regression. For these data, beta coefficients with corresponding *t* statistics are presented. Categorical data were analyzed by using logistic regression. For these data, odds ratios have been computed based on regression coefficients. CI = confidence interval.

[a]Adjusted R^2 = .027.

[b]Adjusted R^2 = .091.

[c]Adjusted R^2 = .059.

[d]Sign indicates the direction of the relationship between dependent and independent variables.

*$P \leq .05$.

of psychiatrists. The size of the hospitals' service area (as defined by hospital administrators) was also included because facilities serving a very large geographic area (e.g., a state or region) were presumed to have difficulty providing as much transitional care as those serving smaller areas, regardless of patients' needs or revenue sources. Of these, average daily psychiatric census neared significance as a positive predictor of the provision of case management ($\beta = .0008$, $P = .09$); degree of urbanicity was significant as a positive predictor of hospital responsibility for aftercare ($\beta = .003$, $P = .017$). These two variables were included in further analyses.

Regression Analysis

Regression analyses were used to examine the impact of the independent variables described above on the provision of transitional care. Dependent variables measuring hospital responsibility for aftercare, length of follow-up, and intensity of follow-up are ordinal and were analyzed by using a multivariate least squares regression model. Logistic regression was used in the analysis of categorical variables: assignment of full-time staff to patient follow-up and the provision of case management. For these variables, odds ratios and 95% confidence intervals were calculated based on the regression coefficients. Because almost all facilities surveyed (93%) provided team review of patient discharge plans, this variable was not included in the bivariate or multivariate analyses.

An additional series of regression analyses was done excluding hospitals that reported "never" taking responsibility for ensuring aftercare. We were interested to see if once a hospital has made a commitment to ensuring aftercare, the pattern of factors influencing transitional care might be different from those affecting all hospitals, including those that report never taking responsibility (15% of the current sample). Interestingly, the results were virtually identical to those found for the entire sample.

Results

Descriptive Statistics

Table 2 presents a profile of the kind and quantity of transitional care services available in specialty and general hospitals. Case management, generally regarded as the most clinically effective form of transitional care, was provided by only 26% of respondents. Seventy-three percent reported follow-up of discharged patients for 1 month or less; only half (54%) reported follow-up of more than 1 week. To address overlap between hospitals that reported assigning staff exclusively to patient follow-up and those that reported providing case management, we did additional analyses. These showed that 35% of the sample provided neither case management nor full-time follow-up

staff; 40% had full-time follow-up staff but did not provide case management.

We found that almost all inpatient facilities surveyed (93%) conducted a team review of patients' discharge plans; however, there is a great deal of variation among hospitals in how discharge plans are carried out. Over half (59%) of the facilities had staff assigned exclusively to patient follow-up, but when asked how often the hospital (vs the patient or the physician) took responsibility for ensuring aftercare, 35% answered "sometimes" and 50% answered "often." When asked what method of follow-up was used most often, 78% reported that their hospital either set up initial aftercare appointments for patients or made a staff person available to help with a referral.

Regression

Hospital type (psychiatric vs general) was by far the strongest determinant of transitional care practices (Table 3). After controlling for case mix, among other factors, psychiatric hospitals reported taking more responsibility than did general hospital psychiatric units for ensuring that patients received aftercare ($\beta = -.223$, $P = .0003$). Psychiatric hospitals also provided follow-up for patients for a longer period ($\beta = -.705$, $P = .0001$), and used more intensive methods of follow-up ($\beta = -.693$, $P = .0001$). Specialty hospitals also were more likely to assign full-time staff to follow-up ($\beta = -.809$, $P = .0001$).

The provision of case management was not predicted by hospital type. It was, however, associated positively with severity of patient illness (percentage diagnosed as schizophrenic) ($\beta = .021$, $P = .0001$). The level of state mental health funding (other than Medicaid) ($\beta = .010$, $P = .0005$) was also a positive predictor and may be an index of severity of illness.

Duration of follow-up, although most strongly and positively associated with hospital type, was also negatively predicted by severity of illness ($\beta = -.005$, $P = .041$); that is, the greater the percentage of patients with schizophrenia in a facility, the shorter the average length of follow-up care provided. Offering residency training was positively related to four of the five transitional care indices: length of follow-up ($\beta = .182$, $P = .05$), hospital responsibility for ensuring aftercare ($\beta = .125$, $P = .054$), intensity of follow-up methods ($\beta = .232$, $P = .051$), and provision of case management ($\beta = .511$, $P = .006$). Level of funding from the state was negatively associated with level of hospital responsibility ($\beta = -.003$, $P = .03$).

Discussion

Not surprisingly, the type of hospital (specialty vs general) strongly influences how the institutions' managers perceive their role and affects their willingness to provide transitional care. Specialty hospitals tend to be freestanding hospitals located at a distance from other health care services; nonetheless, they often serve as referral centers and provide both acute and long-term services, as well as more specialized treatment programs. They usually are familiar with the range of services needed by their patients and work with the local service provider network. Although specialty hospitals have varied arrangements for providing ambulatory care, these facilities generally can provide ongoing treatment by their medical staffs or can help negotiate treatment by professionals in the community.

General hospitals, on the other hand, often view psychiatric services as a small part of their overall activity and usually treat only acute cases on relatively small (e.g., 20-bed) units. These units are part of a larger health care institution and community-care system that tends to be located in an urban area. General hospitals, therefore, often depend on local providers, such as community mental health centers, to take over postdischarge treatment. A notable exception to this trend is that 36% of facilities providing case management were nonprofit general hospitals. That could reflect the mainstreaming of the mentally ill to the acute medical care system. In all, these characteristics related to type predominate over ownership form per se in determining the hospital's behavior in providing transitional services. A special case is the public (state or county) mental hospital, which provides primarily intermediate and long-term care; after prolonged efforts by state governments to deinstitutionalize, many now place heavy emphasis on aftercare.

Case management generically may be viewed as the most intensive and assertive form of transitional or bridging services. We find that despite its importance in the eyes of family members, policymakers, managed care companies, and community mental health clinicians, this service is not very widely employed by hospitals—only 26% provide it, perhaps because it was a largely "non-reimbursable" service at the time of the study.

We found that transitional care was most likely to be available when the institution had a high number of patients who were severely ill (e.g., schizophrenia) and that there were strong correlations between this variable and the provision of case management. The reliance on relatively short-term follow-up and little patient contact to establish aftercare (Table 1) suggests that hospitals may recognize the need for follow-up but invest relatively few resources in doing it. Some of the factors associated with more efforts across the board to provide transitional care were the presence of state or county contracts for care, larger units, location in urban areas, and the presence of a residency training program. Apparently, teaching hospitals believe that providing transitional services is an important part of learning about clinical practice and they may allow additional staff time and incentives to follow patients as part of the training experience.

Conclusion

A national survey of mental health facilities found that transitional care services for mentally ill patients leaving the hospital were uneven and disappointing. Although there is widespread recognition of the need for providing transitional care services, there is also wide variability in the extent and willingness of hospitals to take responsibility for, and intensively implement, measures to ensure effective aftercare. Financial barriers appear to play a critical role. Hospitals that have a high proportion of patients with severe and persistent disorders and that receive nonfederal public funding for providing services are more likely than others to provide intensive follow-up, including case management. Although mental health care providers are increasingly influenced by pressures to economize, it is important that they remain aware of the need for structured, intensive, and specific transitional care interventions for all psychiatric patients. Private as well as public funding sources must recognize that appropriately designed incentives can play a role in encouraging institutions to provide this type of service. For public facilities, better transitional and aftercare services can reduce readmission rates and reduce costs. For private hospitals, changes in the financing of mental health care toward network and capitation arrangements under managed care systems should create new incentives to do likewise. To ensure adequate aftercare and case-management

services for severely mentally ill persons, national health care reform proposals should either provide for reimbursement (in the manner of current Medicaid programs) or else explicitly mandate responsibility for such services to state mental health authorities.[25]

Acknowledgments

The authors are grateful for comments and statistical review given by Sherrie Epstein, William Fisher, and Mark Schlesinger.

References

1. Fisher WH, Dorwart RA, Schlesinger M, Epstein S, Davidson H. The privatization of inpatient treatment for the seriously mentally ill: assessing the roles of public and private general hospitals. *Hosp Community Psychiatry*. In press.
2. Bellack AS, Mueser KT. A comprehensive treatment program for schizophrenia and chronic mental illness. *Community Mental Health J*. 1986;22:175–189.
3. Fisher WH, Geller JL, Altaffer F, Bennett MB. The relationship between community resources and state recidivism. *Am J Psychiatry*. 1992;149:385–390.
4. Shore MF, Cohen MD. The Robert Wood Johnson Foundation program on chronic mental illness: an overview. *Hosp Community Psychiatry*. 1990;41:1212–1216.
5. Dorwart RA. Managed mental health care: myths and realities in the 1990's. *Hosp Community Psychiatry*. 1990;41:1087–1091.
6. Dorwart RA, Schlesinger M. Privatization of psychiatric services. *Am J Psychiatry*. 1988;145:543–553.
7. Mechanic D. Managed care for the seriously mentally ill. *Am J Public Health*. 1992;82:788–789. Editorial.
8. Fisher WH, Dorwart RA, Schlesinger M, Davidson H. Contracting between public agencies and private psychiatric inpatient facilities. *Med Care*. 1991;29:766–774.
9. Bachrach L. The challenge of service planning for chronic mental patients. *Community Mental Health J*. 1986;22:170–174.
10. Torrey EF. Continuous treatment teams in the care of the chronically mentally ill. *Hosp Community Psychiatry*. 1986;37:1243–1247.
11. Siegel C, Alexander MJ, Lin S. Severe alcoholism in the mental health sector, II: effects of service utilization on readmission. *J Stud Alcohol*. 1984;45:510–516.
12. Solomon P, Gordon B, Davis JM. Reconceptualizing assumptions about community mental health. *Hosp Community Psychiatry*. 1986;37:708–712.
13. Rosenfield S, Canton C, Nachumi G, Robbins E. Closing the gaps: the effectiveness of linking programs connecting chronic mental patients from the hospital to the community. *J Appl Behav Sci*. 1986;22:411–423.
14. Myerson AT, Herman GS. What's new in aftercare? A review of recent literature. *Hosp Community Psychiatry*. 1983;34:333–342.
15. Herman NJ, Smith CM. Mental hospital depopulation in Canada: patient perspectives. *Can J Psychiatry*. 1989;34:386–391.
16. Robinson GK, Toff-Bergman G. *Choices in Case Management: Current Knowledge and Practice for Mental Health Programs*. Washington, DC: Mental Health Policy Resource Center; 1990.
17. Surles RC, Blanch AK, Shern DL, Donahue SA. Case management as a strategy for systems change. *Health Affairs*. Spring 1992;11:151–163.
18. Bachrach LL. Continuity of care and approaches to case management for long-term mentally ill patients. *Hosp Community Psychiatry*. 1993;44:465–468.
19. *Admission and Discharge Practices of Psychiatric Hospitals*. Albany, NY: New York State Commission on Quality of Care for the Mentally Disabled; April 1988.
20. Shortell SM, Morrison EM, Hughes SL, Friedman B, Coverdill J, Berg L. The effects of hospital ownership and the practice of medicine. In: Gray BH, ed. *For-Profit Enterprises in Health Care*. Washington, DC: National Academy Press; 1986:402–421.
21. Clark RE, Fox TS. A framework for evaluating the economic impact of case management. *Hosp Community Psychiatry*. 1993;44:469–473.
22. Dorwart RA, Schlesinger M, Davidson H, Epstein S, Hoover C. A national study of psychiatric hospital care. *Am J Psychiatry*. 1991;148:204–210.
23. Redick RW, Stroup A, Witkin MJ, et al. *Private Psychiatric Hospitals, United States: 1983–84 and 1986*. Rockville, Md: NIMH Division of Biometry and Epidemiology, Survey and Reports Branch; October 1989. Mental Health Statistical Note 191.
24. Office of Data Analysis and Management, Bureau of Health Professions, US Department of Commerce. *User Documentation for the ODAM Area Resource File (ARF)*. Springfield, Va: National Technical Information Service; March 1988.
25. Health Security Act, § 1757, 103d Cong, 1st Sess (1993).

Call for Abstracts for Injury Control Late-Breaker Session

The Injury Control and Emergency Health Services Section of the American Public Health Association will again feature a late-breaker session during the APHA annual meeting October 30 through November 3, 1994, in Washington, DC. The session will be held on Thursday, November 3, at 8:30 AM and will feature work completed within the last few months, after the deadline for consideration in the regular symposia of the meeting.

Abstracts of 250 words or fewer will be accepted by the section until *September 10, 1994*. No special form is required. Please send or fax the abstract, title of paper, author's name, address, telephone, and fax number to Rick Waxweiler, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, 4770 Buford Hwy, NE, Mail Stop F41, Chamblee, GA 30341; tel (404) 488-4031; fax (404) 488-4338. All submitters will be notified of decision by fax by September 24.

What Determines Mortality Risk in Male Former Cigarette Smokers?

ABSTRACT

Objectives. The purpose of this study was to examine what factors determine the mortality experience of male ex-cigarette smokers, those who no longer smoke at all and those who changed to pipe or cigar smoking.

Methods. A cohort study was undertaken with 18-year mortality data on 19 018 men.

Results. Ex-cigarette smokers had an intermediate mortality risk compared with never and current smokers. Ex-cigarette smokers who switched to pipe smoking had higher mortality than those who no longer smoked at all. The mortality rates for pipe and cigar smokers who were former cigarette smokers were higher than those for pipe or cigar smokers who had never smoked cigarettes. Ex-cigarette smokers who consumed more than 20 cigarettes per day for more than 20 years experienced increased mortality for both coronary heart disease and neoplasms, even after 30 years of cessation.

Conclusions. These results support the notion that an elevated mortality risk may be seen for ex-cigarette smokers, even after they have given up smoking for many years. Ex-cigarette smokers who change to a pipe have a greater mortality risk than those who no longer smoke at all. (*Am J Public Health.* 1994;84:1235–1242)

Yoav Ben-Shlomo, MRCP, George Davey Smith, MD, Martin J. Shipley, MSc, and Michael G. Marmot, FFPHM

Introduction

The prevalence of smoking in many populations has been falling, in particular for men.[1] An ever-increasing proportion of the general population are former cigarette smokers who have either given up smoking or changed to a pipe or cigars. Many studies have reported reduced mortality for former smokers compared with smokers.[2–9] Whether former smokers ever reach the same levels of mortality risk as never smokers remains controversial.[10–12] The mortality risk of pipe or cigar smokers may also differ depending on a past history of cigarette smoking. These issues are of importance because they may provide some insight as to etiological mechanisms, enable better predictions of future trends in mortality, and provide the public with realistic expectations of the potential benefit, in terms of the reduced risk of mortality, that has been observed from stopping smoking.

The Whitehall Study previously reported no difference in risk of death from coronary heart disease between never smokers and ex-smokers after 5 years of follow-up.[13] We re-examined data from the Whitehall study after 18 years of follow-up. Our aims were to examine the mortality risk of former cigarette smokers compared with never smokers with particular reference to duration of smoking habit, number of cigarettes consumed, change in smoking habit, and time since giving up cigarettes. We also looked at the effect of changing from cigarettes to pipe or cigar smoking.

Subjects and Methods

In the Whitehall Study, 19 018 men aged 40–69 years from the British Civil Service were examined between 1967 and 1969. The main aim of the study was to examine what risk factors may be important in the etiology of coronary heart disease, as well as other chronic diseases.[14–16] In this cohort study, all exposure data were collected only at baseline with no subsequent reassessments.

If the subjects stated at baseline that they had never smoked any form of tobacco they were classified as never smokers. If they currently smoked cigarettes they were categorized as current cigarette smokers, irrespective of whether they also smoked a pipe or cigars. If subjects did not smoke cigarettes or any other form of tobacco at baseline but had ever smoked cigarettes, they were categorized as ex-cigarette smokers. Subjects who were ex-cigarette smokers but current pipe or cigar smokers were referred to as secondary pipe or cigar smokers. Subjects who had never smoked cigarettes but were current pipe or cigar smokers were referred to as primary pipe or cigar smokers. Current smokers of both pipe and cigars were excluded so that the effects of either pipe or cigar smoking individually could be examined. Three hundred ninety-five men who were former pipe or cigar smokers and no longer smoked at all were omitted from further analysis. For both current and ex-smokers, the age when they started to

Yoav Ben-Shlomo, Martin J. Shipley, and Michael G. Marmot are with the Department of Epidemiology and Public Health, University College London, England. George Davey Smith is with the Department of Public Health, University of Glasgow, Glasgow, Scotland.

Requests for reprints should be sent to Yoav Ben-Shlomo, MRCP, Department of Epidemiology and Public Health, University College London, 66-72 Gower St, London WC1E 6EA United Kingdom.

This paper was accepted November 29, 1993.

TABLE 1—Distribution of Subjects, by Smoking Status

Smoking Status	No. (%) of Subjects
Total sample size	19 018
No data on smoking status	14
Both pipe and cigar smokers (excluded from analysis)	239
Ex–pipe or ex–cigar smokers (excluded from analysis)	395
Remaining sample size	18 370 (100)
Never smokers	3 105 (16.9)
Ex–cigarette smokers	4 833 (26.3)
Current cigarette smokers	7 921 (43.1)
Pipe smokers	1 748 (9.5)
Primary	492 (28.1)
Secondary	1 256 (71.9)
Cigar smokers	763 (4.2)
Primary	105 (13.8)
Secondary	658 (86.2)

smoke and the most they ever smoked regularly for as long as a year were ascertained. A small subsample of current cigarette smokers were later recruited into an intervention trial of smoking advice.[17]

Additional questions were asked about chest pain, respiratory symptoms, and civil service employment grade. Civil service employment grade is in four levels: administrators, professionals and executives, clerical, and "other" (mainly unskilled manual) grades. Administrators are the top grades, with the highest status and income in the civil service; the "other" grades are the lowest. For 873 subjects employed by the diplomatic service and the British Council, grades were not comparable to those for the rest of the sample. For analyses that involved grade, these subjects were kept as a separate group.

Clinical measurements included height, weight, blood pressure, forced expiratory volume in 1 second ($FEV_{1.0}$), forced vital capacity, and a limb lead electrocardiogram. An oral glucose tolerance test was administered (50 g) after an overnight fast, and a 2-hour capillary blood sample was drawn for the measurement of glucose and cholesterol concentrations.

Disease at entry was defined by any of the following: relative shortness of breath on level ground, probable or possible intermittent claudication,[18] past history of diabetes, heart or blood pressure trouble, unexplained weight loss over the preceding year, grade 1 or 2 angina according to the Rose angina questionnaire,[14] severe chest pain for over half an hour, and an abnormal resting electrocardiogram according to the following Minnesota code items: Q/QS waves (1.1–1.3); ST depressions (4.1–4.4); T-wave inversion or flattening (5.1–5.3); or left-bundle branch block (7.1). Full details of the procedures used were previously reported.[19]

Records from over 99% of subjects were flagged at the National Health Service Central Registry. Death certificates were coded according to the eighth revision of the *International Classification of Diseases* (ICD), and this almost-complete mortality follow-up to January 31, 1987, provides the basis for these analyses. Death was classified as being due to coronary heart disease (ICD codes 410–414); cardiovascular disease (ICD codes 390–458); neoplasm of trachea, bronchus, and lung (ICD code 162)—henceforth referred to as lung cancer—or any neoplasm (ICD codes 140–239); and respiratory disease (ICD codes 460–519). Neoplasms were also classified according to whether smoking is considered to play a role in their etiology.[2,5,20] The causes deemed to be smoking related are malignant neoplasms of the lip (ICD code 140), tongue (ICD code 141), mouth and pharynx (ICD codes 143–149), esophagus (ICD code 150), and pancreas (ICD code 157); respiratory system neoplasms (ICD codes 160–163) and urinary system neoplasms (ICD codes 188–189); and malignant neoplasms of unspecified site or secondary neoplasms (ICD codes 195–199). All other cancers were classified as not related to smoking.

The number of years since stopping smoking is often taken from the date of quitting up to the start of the study. That is reasonable when the period of follow-up is relatively short. With 18 years of follow-up, however, it will result in misclassification, as the number of years since stopping smoking for an ex-smoker actually increases over the follow-up period. To allow for this time dependence, the time data were analyzed by using the standard person-years approach.[21] For example, a subject aged 48 years at the start of the study who died after 13.5 years of follow-up would contribute 2 person-years to the age category 45–49 years, 5 person-years to the age categories of 50–54 and 55–59 years, and 1.5 person-years to the age category of 60–64 years. His death would be allocated to the 60–64-year category. If in fact this subject was an ex–cigarette smoker who had given up 3 years before the start of the study, then he would contribute 7 person-years to the 1–9 years since quitting cigarettes category and 6.5 person-years to the 10–19 years since quitting category, with his death being allocated to this latter category. Because of the size of the data set, person-years of observation during the 18-year follow-up were allocated on a year-by-year basis. For each year of follow-up, a new record was created consisting of each subject's current age together with baseline characteristics and length of follow-up during that year. For ex-smokers, the time since stopping smoking up to that year of follow-up was calculated as in the example above. This expanded data set was then used as the basis for all analyses of mortality.

Mortality rates were calculated by using the person-years at risk as outlined above. These rates and also all means and proportions were standardized for age by the direct method, with the total population as the standard. Differences in proportions and means in relation to smoking status were tested with the Cochran-Mantel-Haenszel statistic in SAS[22] and by analysis of covariance, respectively. Adjustment for risk factors and calculation of confidence intervals (CIs) for rate ratios were done by using Cox's proportional hazards regression model.[23] These methods are very similar to those that use Poisson regression, but they do not require the assumption of a constant hazard within the individual person-years of follow-up. Dummy variables were used in the models to assess the effects of various risk factor categories (e.g., duration of smoking, maximum amount smoked, and time since quitting) in ex–cigarette smokers compared with never smokers. Tests for linear trend in these three risk factors among ex–cigarette smokers were computed by including a dummy variable for ex–cigarette smoker and a continuous variable for the relevant risk factor. The coefficient for the continuous variable therefore represents the change in mortality per unit change in the risk factor among ex-smokers and was used as a test for trend in the risk factor. Significant departures from linearity were examined by calculating the difference for the chi-square value for heterogeneity and the chi-square value for linear trend.

TABLE 2—Age-Adjusted Mortality Rates per 1000 Person-Years, by Smoking Status

Smoking Status	Mortality Rate per 1000 Person-Years (No. of Deaths)							
				Neoplasms				
	All Causes	CHD	CVD	Total	SRN[a]	NSRN[b]	Lung Cancer	Respiratory Disease
Never smoker (n = 3105)	10.50 (421)	4.34 (175)	6.06 (240)	2.71 (111)	0.96 (49)	1.75 (72)	0.25 (24)	0.68 (24)
Ex-cigarette smoker (n = 4833)	12.10 (969)	4.72 (379)	6.48 (519)	3.74 (300)	1.51 (121)	2.23 (179)	0.73 (58)	0.95 (75)
Primary pipe (n = 492)	13.94 (113)	5.64 (46)	7.85 (64)	3.37 (27)	1.64 (13)	1.73 (14)	1.00 (8)	1.37 (11)
Secondary pipe (n = 1256)	14.12 (282)	5.92 (120)	8.13 (161)	4.20 (85)	2.43 (49)	1.77 (36)	1.18 (24)	1.00 (20)
Primary cigar (n = 105)	5.04 (9)	1.96 (4)	2.36 (5)	2.48 (4)	0.85 (2)	1.63 (2)	0.45 (1)	0.00 (0)
Secondary cigar (n = 658)	12.64 (132)	3.95 (42)	6.06 (64)	4.15 (43)	2.38 (25)	1.77 (18)	1.91 (20)	0.97 (10)
Current cigarette (n = 7921)	20.95 (2570)	7.56 (929)	10.52 (1292)	6.64 (816)	4.28 (524)	2.37 (292)	2.98 (365)	2.20 (266)

Note. CHD = coronary heart disease; CVD = cardiovascular disease; SRN = smoking-related neoplasm; NSRN = non-smoking-related neoplasm.
[a]Smoking-related neoplasms are malignant neoplasms of the lip (ICD code 140), tongue (ICD code 141), mouth and pharynx (ICD codes 143–149), oesophagus (ICD code 150), and pancreas (ICD code 157); respiratory system neoplasms (ICD codes 160–163) and urinary system neoplasms (ICD codes 188–189); and malignant neoplasms of unspecified site or secondary neoplasms (ICD codes 195–199).
[b]Non-smoking-related neoplasms were all other cancers not classified as smoking related.

For tabulations of smoking status with other baseline characteristics, the number of years since quitting at entry to the study was used instead.

Results

The different smoking categories are shown in Table 1. Of the 7921 (43.1%) men classified as cigarette smokers, 1363 (17.2%) also smoked a pipe and/or cigars. A total of 714 (9.0%) men were recruited into an intervention study and given antismoking advice. The age-adjusted mortality rates for these smoking-status groups are shown in Table 2 by cause or category of death. Current smokers had the highest mortality rates for each group of conditions. Mortality rates for ex–cigarette smokers were between those of never smokers and current smokers, although they usually were closer to the rates for never smokers. Both primary and secondary pipe smokers had significantly elevated mortality rate ratios, around 40% above those of never smokers, for all causes and coronary heart disease. Secondary pipe or cigar smokers had higher mortality compared with primary pipe or cigar smokers, but only all-cause mortality for secondary compared with primary cigar smokers was statistically significant (rate ratio = 2.30; 95% CI = 1.17, 4.52). Secondary pipe smokers did have higher mortality than ex–cigarette smokers for all-cause mortality (rate ratio = 1.17; 95% CI = 1.03, 1.34), coronary heart disease (rate ratio = 1.27; 95% CI = 1.03, 1.56), cardiovascular disease (rate ratio = 1.25; 95% CI = 1.05, 1.49), and smoking-

TABLE 3—Association between Smoking Habit, Years since Giving Up at Baseline, and Prevalence (%) of Preexisting Disease at the Initial Survey, Adjusted for Age

Smoking Status	Percentage of Subjects with Preexisting Disease (SE)			
	Doctor Diagnosis of High Blood Pressure or Heart Disease	Electrocardiogram Ischemia[a]	Dyspnea	Disease at Entry[a]
Never smoker	4.08 (0.40)	14.5* (0.7)	3.98** (0.39)	21.0* (0.8)
Ex–cigarette smoker	4.69 (0.30)	16.2 (0.5)	5.48 (0.32)	23.5 (0.6)
> 20 y[b]	4.67 (0.61)	16.5 (1.1)	4.93 (0.69)	23.9 (1.3)
16–20 y	5.10 (0.86)	16.1 (1.4)	4.52 (0.82)	24.1 (1.7)
11–15 y	3.66 (0.68)	13.6 (1.2)	4.58 (0.76)	19.8 (1.4)
6–10 y	4.75 (0.70)	16.3 (1.2)	4.89 (0.72)	22.5 (1.4)
0–5 y	4.93 (0.58)	17.1 (1.0)	7.50 (0.70)	25.2 (1.2)
P for trend test among ex-smokers	.62	.79	< .001	.86
P for departures from trend test among ex-smokers	.50	.17	.09	.03
Current smoker	3.73** (0.21)	16.6 (0.4)	7.00*** (0.28)	25.2* (0.5)

[a]See the text for a detailed definition of this condition.
[b]Years since quitting reported at baseline, not including any follow-up period.
*$P < .05$; **$P < .01$; ***$P < .001$ for tests of differences in prevalence for never or current smokers compared with all ex-smokers.

TABLE 4—Number of Deaths and Rate Ratio[a] (95% CI) for Ex–Cigarette Smokers Compared with Never Smokers, by Amount Smoked, Duration of Smoking in Years, and Period since Quitting Cigarettes

Cause of Death	Never Smoker	Duration of Smoking, y				Rate Ratio[b] per 10 y	Maximum No. of Cigarettes/d Consumed				Rate Ratio[b] per 10 cigarettes/d	Years since Quitting				Rate Ratio[b] per 10 y
		0–9	10–19	20–29	≥30		1–9	10–19	20–29	≥30		1–9	10–19	20–29	≥30	
All causes	421	114	158	307	382		165	309	313	172		103	299	274	278	
	1.00	1.02 (0.83, 1.25)	0.94 (0.78, 1.13)	1.29 (1.11, 1.49)	1.44 (1.25, 1.65)	1.15*** (1.09, 1.21)	1.00 (0.84, 1.20)	1.14 (0.98, 1.32)	1.32 (1.14, 1.53)	1.48 (1.24, 1.76)	1.11*** (1.06, 1.16)	1.59 (1.28, 1.98)	1.33 (1.15, 1.54)	1.13 (0.97, 1.32)	1.11 (0.95, 2.34)	0.91** (0.86, 0.96)
CHD	175	50	63	119	145		70	114	127	64		41	113	111	113	
	1.00	1.07 (0.78, 1.46)	0.90 (0.67, 1.19)	1.41 (1.13, 1.76)	1.90 (1.62, 2.24)	1.14** (1.05, 1.25)	1.05 (0.79, 1.38)	1.03 (0.82, 1.31)	1.32 (1.05, 1.66)	1.37 (1.03, 1.83)	1.12** (1.03, 1.21)	1.44 (1.03, 2.03)	1.19 (0.94, 1.50)	1.13 (0.89, 1.43)	1.15 (0.91, 1.46)	0.97 (0.89, 1.06)
Neoplasms																
Total	111	29	52	106	112		37	98	105	58		42	92	86	79	
	1.00	0.98 (0.65, 1.47)	1.17 (0.84, 1.63)	1.70 (1.30, 2.21)	1.66 (1.28, 2.17)	1.17** (1.06, 1.29)	0.87 (0.60, 1.26)	1.39 (1.06, 1.83)	1.71 (1.31, 2.23)	1.93 (1.40, 2.66)	1.16*** (1.06, 1.26)	2.39 (1.67, 3.42)	1.51 (1.15, 2.00)	1.37 (1.03, 1.81)	1.23 (0.92, 1.65)	0.84** (0.76, 0.94)
SRN	39	9	17	43	52		8	33	50	29		20	42	30	29	
	1.00	0.87 (0.42, 1.79)	1.10 (0.62, 1.95)	1.94 (1.26, 2.99)	2.09 (1.37, 3.17)	1.29** (1.10, 1.51)	0.53 (0.25, 1.13)	1.30 (0.82, 2.08)	2.28 (1.50, 3.47)	2.67 (1.65, 4.32)	1.30*** (1.16, 1.46)	3.15 (1.86, 5.33)	1.89 (1.25, 2.85)	1.31 (0.83, 2.06)	1.27 (0.75, 2.13)	0.76** (0.65, 0.90)
NSRN	72	20	35	63	60		29	65	55	29		22	50	56	50	
	1.00	1.03 (0.63, 1.70)	1.20 (0.80, 1.79)	1.58 (1.12, 2.21)	1.47 (1.04, 2.09)	1.12 (0.98, 1.27)	1.06 (0.69, 1.63)	1.46 (1.05, 2.05)	1.40 (0.99, 1.99)	1.54 (1.00, 2.38)	1.04 (0.93, 1.18)	1.94 (1.21, 3.12)	1.29 (0.91, 1.83)	1.39 (0.99, 1.95)	1.35 (0.92, 1.99)	0.91 (0.79, 1.04)
Lung cancer	10	2	5	18	33		3	10	26	19		14	23	15	6	
	1.00	0.75 (0.16, 3.41)	1.27 (0.74, 2.20)	3.14 (1.45, 6.79)	5.05 (2.48, 10.3)	1.66*** (1.31, 2.10)	0.77 (0.21, 2.81)	1.51 (0.63, 3.62)	4.58 (2.21, 9.51)	6.70 (3.11, 14.4)	1.40*** (1.20, 1.62)	8.68 (4.00, 18.9)	4.08 (2.03, 8.20)	2.59 (1.21, 5.54)	1.00 (0.32, 3.10)	0.53*** (0.40, 0.70)

Note. CHD = coronary heart disease; SRN = smoking-related neoplasm; NSRN = non-smoking-related neoplasm.

[a]Rate ratios are adjusted for age and grade.

[b]Change in mortality (rate ratio) for a given change in the risk factor among ex-smokers.

*$P < .05$; **$P < .01$; ***$P < .001$.

TABLE 5—Effects on Mortality Risk of the Duration of Smoking and the Amount Smoked, for Ex–Cigarette Smokers and Current Smokers Compared with Never Smokers

Duration of Smoking/ Maximum No. of Cigarettes per Day	Ex-Smokers, by Years Since Quitting									Current Smokers	
	≥30 y			10–29 y			1–9 y				
	No. of Deaths	RR	(95% CI)	No. of Deaths	RR	(95% CI)	No. of Deaths	RR	(95% CI)	RR	(95% CI)
All causes of death											
≤19 y											
1–19 cigs/d	123	0.91	(0.74, 1.11)	42	0.77	(0.56, 1.05)	1[a]	...	...		
≥20 cigs/d	66	1.34	(1.04, 1.74)	35	0.99	(0.70, 1.40)	0[a]	...	...		
≥20 y											
1–19 cigs/d	36	1.32	(0.94, 1.86)	225	1.24	(1.06, 1.46)	43	1.49	(1.09, 2.04)		
≥20 cigs/d	47	1.42	(1.05, 1.92)	279	1.41	(1.21, 1.64)	58	1.69	(1.28, 2.22)		
Current smoking										2.13	(1.92, 2.36)
Coronary heart disease											
≤19 y											
1–19 cigs/d	56	1.04	(0.77, 1.40)	15	0.60	(0.35, 1.01)	0[a]	...	...		
≥20 cigs/d	26	1.31	(0.87, 1.99)	12	0.74	(0.41, 1.33)	0[a]	...	...		
≥20 y											
1–19 cigs/d	9	0.92	(0.47, 1.81)	85	1.18	(0.91, 1.53)	17	1.34	(0.81, 2.21)		
≥20 cigs/d	18	1.52	(0.93, 2.47)	111	1.41	(1.11, 1.79)	24	1.60	(1.04, 2.45)		
Current smoking										1.90	(1.62, 2.24)
All neoplasms											
≤19 y											
1–19 cigs/d	28	0.80	(0.53, 1.21)	12	0.79	(0.44, 1.44)	1[a]	...	...		
≥20 cigs/d	24	1.90	(1.22, 2.95)	15	1.53	(0.89, 2.64)	0[a]	...	...		
≥20 y											
1–19 cigs/d	13	1.95	(1.09, 3.49)	62	1.33	(0.98, 1.82)	18	2.30	(1.40, 3.80)		
≥20 cigs/d	13	1.61	(0.90, 2.87)	89	1.75	(1.32, 2.31)	22	2.37	(1.50, 3.75)		
Current smoking										2.61	(2.14, 3.18)
Smoking-related neoplasms											
≤19 y											
1–19 cigs/d	11	0.90	(0.45, 1.77)	2	0.40	(0.10, 1.64)	0[a]	...	...		
≥20 cigs/d	6	1.35	(0.57, 3.18)	6	1.86	(0.78, 4.40)	0[a]	...	...		
≥20 y											
1–19 cigs/d	5	1.97	(0.77, 5.03)	18	1.08	(0.61, 1.88)	5	1.82	(0.72, 4.63)		
≥20 cigs/d	6	1.97	(0.83, 4.68)	46	2.53	(1.65, 3.88)	15	4.61	(2.54, 8.39)		
Current smoking										4.61	(3.32, 6.39)

Note. RR = rate ratio (adjusted for age and employment grade); CI = confidence interval. The rate ratio for never smokers is 1.0.
[a]Rate ratios not calculated because of small numbers. The expected mortality risk for death from all causes, compared with the risk of never smokers, is less than 1.22.

related neoplasm (rate ratio = 1.63; 95% CI = 1.19, 1.49). This difference could not be explained by their previous smoking habits, as cigarette smokers who switched to a pipe consumed significantly fewer cigarettes, smoked for a shorter period, and had given up smoking for longer than cigarette smokers who gave up all forms of tobacco. Multivariate analysis, controlling for these three factors, increased the relative risks for secondary pipe smokers compared with ex–cigarette smokers. No significant differences were found between secondary cigar smokers and ex-cigarette smokers.

We examined the relationship between smoking status and other potentially confounding variables. Current smokers were more likely to be of lower employment grade, had lower systolic blood pressure, and weighed less than never smokers ($P < .001$). No obvious trend was seen for risk factors between never, ex-, and current cigarette smokers except for grade. The proportion of higher grades (administrative and professional/executive) progressively increased with years since giving up smoking (test for linear trend, $P < .001$). Current pipe smokers had similar profiles regardless of whether they had smoked cigarettes in the past. Cigar smokers who had never smoked cigarettes had a more favorable risk factor profile except for being shorter in stature (data available on request).

Because smokers may decide to give up their habit because of preexisting disease, we examined the association between smoking habit and four proxy variables suggestive of preexisting cardiovascular, respiratory, and general disease (Table 3). Former smokers were more likely than current smokers to have a doctor diagnosis of heart disease and hypertension, although the prevalence of

electrocardiogram ischemia was similar for both current and former smokers and greater than that for never smokers. The proportions of reported dyspnea and disease at entry were between those of never smokers and current smokers. In general, smokers who had given up smoking within the last 5 years had higher levels of disease or symptoms than longer-term former smokers.

In view of the inconsistent risk factor profile seen with ex-cigarette smokers, we controlled only for age and grade as potential confounders in all subsequent analyses.

Table 4 shows the age- and grade-adjusted rate ratios for ex-cigarette smokers compared with never smokers by (1) the number of years smoked before giving up, (2) the maximum number of cigarettes consumed for a period of at least 1 year, and (3) the number of years since quitting. Among ex-cigarette smokers, a significant increase in mortality compared with that of never smokers was seen only after 20 years of smoking for all-cause mortality; however, the test for trend for duration smoked was significant for all categories of death except non-smoking-related deaths. Significant increases in mortality were seen as the amount consumed increased, except for non-smoking-related neoplasms. Mortality risk fell significantly with years since quitting for all categories except coronary heart disease and non-smoking-related neoplasms. These two categories showed a marked fall after 10 years, but after that period the risks appeared relatively unchanged.

In this cohort, most of the men started to smoke at a similar age. Therefore, among ex-cigarette smokers, for any given age, duration of smoking will determine the number of years since quitting. Hence, these two variables will be highly (negatively) correlated. They will also be correlated with age. In addition, the maximum amount smoked in any year is likely to be higher in those men who smoked the longest. We did not attempt to control for these highly correlated variables simultaneously in a multivariate model because that would require us to make the strong assumption that the effect of each of these factors on the risk of mortality was multiplicative. Instead, to examine the possible effects of the above factors, we carried out a stratified analysis and fitted models by using a dummy variable for each stratum. That allowed the joint effects of the three factors of interest to be assessed relative to men who had never smoked while controlling for the confounding effects of age and grade (Table 5).

Ex-smokers who never smoked more than 20 cigarettes a day and smoked for less than 20 years showed no evidence of any increased mortality risk for either all-cause mortality or other categories. All-cause mortality remained elevated for heavy smokers who smoked for more than 20 years even after 30 years since quitting. Different patterns emerged for coronary heart disease and neoplasms. Coronary heart disease mortality for former persistent heavy smokers showed an initial fall but then remained at a fairly constant level, although after 30 years this risk was no longer statistically significant. For all neoplasm mortality and smoking-related neoplasm mortality, persistent heavy smokers showed a fall in risk with years since quitting.

Discussion

All-Cause Mortality

Former cigarette smokers had mortality risks between those of current cigarette smokers and those of never smokers, a finding similar to those of other cohort studies.[2,3,7,9] Mortality increased with duration of smoking and maximum amount consumed and decreased with years since quitting. After stratification for these three factors, both the maximum amount smoked and the duration of smoking were independently important in determining mortality risk. Although the risk for ex-smokers fell substantially compared with that for current smokers, persistent heavy smokers showed little further decline with increasing years since quitting. Light smokers who had smoked for less than 20 years had no evidence of any increased risk.

Coronary Heart Disease Mortality

Ex-cigarette smokers showed increased mortality from coronary heart disease compared with never smokers. This mortality risk increased with amount consumed and duration of smoking. A sharp fall in mortality was seen in the first 19 years after quitting. However, after this period there appeared to be little further reduction in risk, and there was no significant trend with years since quitting. Persistent heavy smokers showed a significant elevated risk up to 30 years after quitting.

Several studies have suggested that a risk for coronary heart disease mortality may persist for ex-smokers; however, this risk was small and not always statistically significant.[4,5,7,8,10] Prospective studies that have examined the effects of years since quitting have noted either a sudden fall in risk with little further decline[7,24] or a slower continual decline.[2–4,10] In general, cohort studies support our finding,[2,3,10] whereas case–control studies suggest that the risk for coronary heart disease mortality or nonfatal infarcts falls quickly with no persistent difference between ex–cigarette smokers and never smokers.[11,12] Both the latter studies have had insufficient power to detect a small excess risk.

Fibrinogen has been shown to be elevated in smokers[25] and is a strong predictor of coronary heart disease.[26,27] The levels for ex–cigarette smokers revert to those of never smokers after several years of cessation.[11,25,28] It is an acute and reversible mechanism by which coronary heart disease risk may change. But persistent heavy smokers may have developed irreversible changes in the coronary architecture, unrelated to fibrinogen. A study of postinfarction angina noted an acute reduction in angina for ex-smokers compared with those who continued to smoke, but no difference by 6 years.[29] The role of a reversible, as well as an irreversible, component may also explain the relatively modest reductions for coronary heart disease mortality seen in intervention trials.[9,30,31]

Neoplasm Mortality

Ex–cigarette smokers showed an elevated risk for neoplasm mortality, lung cancer in particular. Significant trends were seen only for smoking-related neoplasms and lung cancer with amount smoked and duration of smoking. Persistent heavy smokers showed a gradual decline in mortality risk with years since quitting, with the risk after 30 years still raised but no longer statistically significant. A persistent elevated risk for both lung cancer and smoking-related cancers has been seen with both mortality data[4,5,7] and incidence data.[32] Intervention studies, including a subsample of subjects recruited from this cohort, also failed to show a reduction of risk from lung cancer.[9,30] A smoker inhales substances that may act as initiators and promoters of carcinogenesis.[33] Smoking cessation may prevent new lesions from developing but may only retard preexisting early lesions, which continue to develop at a slower rate.

Pipe and Cigar Smokers

Many cohort studies have reported the combined mortality rates of cigar and pipe smokers or have failed to distinguish primary from secondary pipe or cigar smoking.[2,3,5,6,8,10,24,34,35] This distinction is important because smokers from older cohorts are more likely to be primary pipe or cigar smokers, whereas recent cohorts will have more secondary pipe or cigar smokers. A recent study reported that 90% of the cigar smokers were ex-cigarette smokers.[36] The small group of primary cigar smokers showed no evidence of an increased mortality risk compared with never smokers. Both primary and secondary pipe smokers were at greater risk of death from coronary heart disease than cigar smokers. This result contradicts some studies[2-4,37] but is supported by another.[8] Both cigar smoking and pipe smoking have been associated with either a greater risk of coronary heart disease mortality[36-38] or no excess risk.[6,24] These inconsistencies may reflect true differences between populations in the type and amount of tobacco consumed, but more likely they reflect confounding by socioeconomic status or other factors, which will differ for different time periods and countries.

Secondary pipe and cigar smokers had a slightly greater risk of death from coronary heart disease than primary pipe or cigar smokers, as has been noted elsewhere,[10] although this difference was not statistically significant. Higher carboxyhemoglobin, higher serum cotinine levels, and an increase in fibrinogen levels have been found in secondary compared with primary cigar smokers.[25,39] Secondary pipe or cigar smokers also had greater risks of smoking-related neoplasm death compared with pure pipe or cigar smokers. Although that could have occurred by chance, it is interesting that the rates of non-smoking-related cancers for primary and secondary pipe and cigar smokers were almost identical. This increased risk may reflect a residual carcinogenic effect from cigarettes. In addition, when cigarette smokers change their habit to pipe or cigars they consume more tobacco and inhale to a greater degree.[40] Changing to a pipe was associated with significantly greater mortality than total cessation of smoking, but no real difference was seen for subjects who took up cigar smoking.

With any observational study, it is unclear whether the benefits seen for smoking cessation are due to smoking cessation itself or to other alterations in life circumstances that may occur concurrently. Baseline differences in other risk factors for this cohort showed little consistent pattern apart from socioeconomic status, as in other studies.[41-46] Ex-smokers may pursue a more healthy life-style, but this is probably a more recent phenomenon.[45]

Smoking status was measured only at baseline and may have been misclassified owing to recidivism among ex-smokers. Empirical data from other studies[9,47] suggest that recidivism usually occurs within the first few years of stopping. Heavy smokers are usually,[9,48] but not always,[47] reported as more likely to resume smoking. Recidivism will in general result in an underestimate of the benefits of giving up; however, in the group of heavy smokers who had given up for more than 30 years, this is unlikely to be a serious bias.

Some smokers are likely to give up because of newly diagnosed disease. This is supported by the observation that ex-smokers who had given up smoking for less than 5 years were more likely than current smokers to have a diagnosis of heart disease or hypertension and to complain of dyspnea. These differences were small and are unlikely to explain any persistent risk among ex-smokers who have given up for more than 30 years, although they may result in underestimating the short-term benefits of giving up.

Conclusions

Ex-smokers show a reduced mortality risk compared with current smokers. This risk may remain persistently higher than the risk for never smokers, depending on previous smoking history and the number of years since quitting. Duration of smoking[10] or consumption of cigarettes[11] alone does not explain the different mortality patterns. Smokers who change to a pipe or cigars appear to retain a greater risk than pipe or cigar smokers who have never consumed cigarettes. Changing to a pipe was associated with a higher mortality risk than that for ex-cigarette smokers who no longer smoked any form of tobacco, but no difference was found for subjects who switched to cigars.

These results have implications for any future predictions about mortality trends for smoking-related diseases. The mortality risk of nonsmokers will differ depending on the proportion of ex-cigarette smokers and their past smoking histories. Hence, the true risk for this group will differ both for different countries and over different time periods. Current life expectancy estimates for ex-smokers[49] may not be valid for future generations.

It is important that the public be aware that whereas stopping smoking results in a markedly diminished risk, a residual excess risk may persist even after many years. The suggestion that ex-smokers have no greater risk than never smokers[50,51] may be counterproductive. Smokers may believe that continuing to smoke is not deleterious as long as they eventually give up. The amount, duration, and period since giving up appear to be important in determining mortality risk.

References

1. Pierce JP. International comparisons of trends in cigarette smoking prevalence. *Am J Public Health*. 1989;79:152–157.
2. Hammond EC. Smoking in relation to the death rates of one million men and women. In: Haenzel W, ed. *Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases*. Bethesda, Md: National Cancer Institute; 1966:127–204. National Cancer Institute Monograph No. 19B.
3. Haramond EC, Horn D. Smoking and death rates—report on forty-four months of follow-up of 187 783 men. II: death rates by cause. *JAMA*. 1958;166:1294–1308.
4. Rogot E, Murray JL. Smoking and causes of death among U.S. veterans: 16 years of observation. *Public Health Rep*. 1980;95:213–221.
5. Doll R, Peto R. Mortality in relation to smoking: 20 years' observations on male doctors. *Br Med J*. 1976;2:1523–1536.
6. Gordon T, Kannel WB, McGee D. Death and coronary attacks in men after giving up cigarette smoking. *Lancet*. 1974;2:1345–1348.
7. LaCroix AZ, Lang J, Scherr P, et al. Smoking and mortality among older men and women in three communities. *N Engl J Med*. 1991;324:1619–1625.
8. Carstensen JM, Pershagen G, Eklund G. Mortality in relation to cigarette and pipe smoking: 16 years' observation of 25 000 Swedish men. *J Epidemiol Community Health*. 1987;41:166–172.
9. Ockene JK, Kuller LH, Svendsen KH, Meilahn E. The relationship of smoking cessation to coronary heart disease and lung cancer in the multiple risk factor intervention trial (MRFIT). *Am. J Public Health*. 1990;80:954–958.
10. Cook DG, Pocock SJ, Shaper AG, Kussick SJ. Giving up smoking and the risk of heart attacks: a report from the British Regional Heart Study. *Lancet*. 1986;2:1376–1380.
11. Dobson AJ, Alexander HM, Heller RF, Lloyd DM. How soon after quitting smoking does risk of heart attack decline? *J Clin Epidemiol*. 1991;44:1247–1253.
12. Rosenberg L, Kaufman DW, Helmrich SP, Shapiro S. The risk of myocardial infarction after quitting smoking in men under 55

years of age. *N Engl J Med.* 1985;313:1511–1514.
13. Reid DD, McCartney PM, Hamilton PJS, Rose G. Smoking and other risk factors for coronary heart-disease in British civil servants. *Lancet.* 1976;2:979–983.
14. Rose G, Hamilton PS, Keen H, Reid DD, McCartney P, Jarrett RJ. Myocardial ischaemia, risk factors and death from coronary heart-disease. *Lancet.* 1977;1:105–109.
15. Fuller JH, Shipley MJ, Rose G, Jarrett RJ, Keen H. Coronary heart disease risk and impaired glucose tolerance: the Whitehall Study. *Lancet.* 1980;1:1373–1376.
16. Marmot MG, Rose G, Shipley M, Hamilton PJS. Employment grade and coronary heart disease in British civil servants. *J Epidemiol Community Health.* 1978;32:244–249.
17. Rose G, Hamilton PJS, Colwell L, Shipley MJ. A randomised controlled trial of anti-smoking advice: 10-year results. *J Epidemiol Community Health.* 1982;36:102–108.
18. Davey Smith G, Shipley MJ, Rose G. Intermittent claudication, heart disease risk factors, and mortality: the Whitehall Study. *Circulation.* 1990;82:1925–1931.
19. Reid DD, Brett GZ, Hamilton PJS, Jarrett RJ, Keen H, Rose G. Cardiorespiratory disease and diabetes among middle-aged male civil servants. *Lancet.* 1974;1:469–473.
20. International Agency for Research on Cancer. *Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Vol 38. Tobacco Smoking.* Lyon, France: International Agency for Research on Cancer; 1986.
21. Berry G. The analysis of mortality by the subject-years method. *Biometrics.* 1983;39:178–184.
22. *SAS User's Guide: Statistics, Version 5.* Cary, NC: SAS Institute Inc; 1985.
23. Cox DR. Regression models and life tables. *J R Stat Soc B.* 1972;34:187–220.
24. Jajich CL, Ostfeld AM, Freeman DH Jr. Smoking and coronary heart disease mortality in the elderly. *JAMA.* 1984;252:2831–2834.
25. Meade TW, Imeson J, Stirling Y. Effects of changes in smoking and other characteristics on clotting factors and the risk of ischaemic heart disease. *Lancet.* 1987;2:986–988.
26. Meade TW, North WRS, Chakrabarti R, et al. Haemostatic function and cardiovascular death: early results of a prospective study. *Lancet.* 1977;1:1050–1054.
27. Yarnell JWG, Baker IA, Sweetnam PM, et al. Fibrinogen, viscosity, and white blood-cell count are major risk factors for ischaemic heart disease—the Caerphilly and Speedwell collaborative heart-disease studies. *Circulation.* 1991;83:836–844.
28. Yarnell JWG, Sweetnam PM, Elwood PC, et al. Some long term effects of smoking on the haemostatic system: a report from the Caerphilly and Speedwell Collaborative Surveys. *J Clin Pathol.* 1987;40:909–913.
29. Daly LE, Graham IM, Hickey N, Mulcahy R. Does stopping smoking delay onset of angina after infarction? *Br Med J.* 1985;291:935–937.
30. Rose G, Colwell L. Randomised controlled trial of anti-smoking advice: final (20 year) results. *J Epidemiol Community Health.* 1992;46:75–77.
31. Holme I, Hjermann I, Helgeland A, Leren P. The Oslo study: diet and antismoking advice. *Prev Med.* 1985;14:279–292.
32. Chyou P, Nomura AMY, Stemmermann GN. A prospective study of the attributable risk of cancer due to cigarette smoking. *Am J Public Health.* 1992;82:37–40.
33. Weisburger JH. The mechanism of lung carcinogenesis and smoking cessation. *Epidemiology.* 1990;1:314–317.
34. Best EWR. *A Canadian Study of Smoking and Health.* Ottawa, Canada: Department of National Health & Welfare; 1966.
35. Gsell OR, Abelin R. Cigar and pipe smoking in relation to lung cancer and excess mortality. *J Natl Cancer Inst.* 1972;48:1795–1803.
36. Kaufman DW, Palmer JR, Rosenberg L, Shapiro S. Cigar and pipe smoking and myocardial infarction in young men. *Br Med J.* 1987;294:1315–1316.
37. Hickey N, Mulcahy R, Daly L, Graham I, O'Donoghue S, Kennedy C. Cigar and pipe smoking related to four year survival of coronary patients. *Br Heart J.* 1983;49:423–426.
38. Gyntelberg F, Lauridsen L, Pedreson PB, Schubell K. Smoking and the risk of myocardial infarction in Copenhagen men aged 40–59 with special reference to cheroot smoking. *Lancet.* 1981;1:987–989.
39. Wald NJ, Idle M, Boreham J, Bailey A. Serum cotinine levels in pipe smokers: evidence against nicotine as cause of coronary heart disease. *Lancet.* 1981;2:775–777.
40. Ockene JK, Pechacek TF, Vogt T, Svendsen K. Does switching from cigarettes to pipes or cigars reduce tobacco smoke exposure? *Am J Public Health.* 1987;77:1412–1416.
41. Friedman GD, Petitti DB, Bawol RD, Siegelaub AB. Mortality in cigarette smokers and quitters. *N Engl J Med.* 1981;304:1407–1410.
42. Gordon T, Kannel WB, Dawber TR, McGee D. Changes associated with quitting cigarette smoking. *Am Heart J.* 1975;90:322–328.
43. Kaprio J, Koskenvuo M. A prospective study of psychological and socioeconomic characteristics, health behaviour and morbidity in cigarette smokers prior to quitting compared to persistent smokers and non-smokers. *J Clin Epidemiol.* 1988;41:139–150.
44. Golbourt U, Medalie JH. Characteristics of smokers, non-smokers and ex-smokers among 10 000 adult males in Israel. *Am J Epidemiol.* 1977;105:75–86.
45. Kato I, Tominaga S, Suzuki T. Characteristics of past smokers. *Int J Epidemiol.* 1989;18:345–354.
46. Kabat GC, Wynder EL. Determinants of quitting smoking. *Am J Public Health.* 1987;77:1301–1305.
47. Sorlie PD, Kannel WB. A description of cigarette smoking cessation and resumption in the Framingham study. *Prev Med.* 1990;19:335–345.
48. Garvey AJ, Bosse R, Glynn RJ, Rosner B. Smoking cessation in a prospective study of healthy adult males: effects of age, time period, and amount smoked. *Am J Public Health.* 1983;73:446–450.
49. Rogers RG, Powell Griner E. Life expectancies of cigarette smokers and nonsmokers in the United States. *Soc Sci Med.* 1991;32:1151–1159.
50. *So You Want to Stop Smoking.* London, England: Health Education Authority; 1991.
51. Fries JF, Green LW, Levine S. Health promotion and the compression of morbidity. *Lancet.* 1989;1:481–483.

The Risk of HIV Infection in a National Sample of Women with Injection Drug–Using Partners

Stephanie Torta, PhD, Mark Beardsley, RhD, Sherry Deren, PhD, and W. Rees Davis, PhD

ABSTRACT

Objectives. This study reports on a large, national cohort of women with injection drug–using sex partners. Information is provided on demographic characteristics; human immunodeficiency virus (HIV) risk factors, including unprotected sex and incidence of sexually transmitted diseases; use of noninjected drugs; HIV serostatus; and other selected health variables.

Methods. A sample of 5162 heterosexual women was recruited for a national acquired immunodeficiency syndrome (AIDS) research and demonstration project. A structured interview was administered, and the women had the option of undergoing HIV testing. Statistical analyses compared three groups on variables of interest: women with single sex partners, women with multiple partners, and women with multiple partners who exchanged sex for drugs and/or money.

Results. These groups differed significantly on virtually all of the demographic and risk variables examined. Women with multiple partners who exchanged sex for drugs and/or money were at higher risk for HIV than women in the other groups, even when selected demographic variables were controlled.

Conclusions. Research is needed on the efficacy of prevention efforts involving these diverse groups of women at risk for AIDS. (*Am J Public Health.* 1994;84:1243–1249)

Introduction

Stimulated by World AIDS Day in December 1990, which highlighted the lack of knowledge regarding HIV in women, efforts have begun to focus on gaining the information necessary to treat and prevent this disease among women.[1] Women with injection drug–using sex partners are among those at highest risk and are the most rapidly growing group of adults with AIDS in the United States.[2,3] As of June 1993, 7495 US women had acquired AIDS as a result of their sexual relationship with an injection drug user.[4] Also, of the 3054 women who contracted AIDS as a result of having sex with an HIV-infected person with risk unspecified, many may have been partners of injection drug users and unaware of their partner's drug use.[5]

Although women with injection drug–using partners often have no history of using injected drugs themselves, many use noninjected drugs, especially crack.[6–8] Interrelationships have been demonstrated among crack use, sexually transmitted diseases, and HIV infection. Crack use is a risk factor for HIV because of its association with high-risk sexual behaviors.[9–11] Cocaine use is also associated with syphilis,[12–14] and sexually transmitted diseases in general are implicated in the enhancement of HIV transmission.[15]

Unfortunately, the number of women who acquire AIDS as a result of sex with infected men is likely to continue to increase.[3] Even when other factors are controlled, male-to-female transmission is more efficient than female-to-male transmission.[16] Also, there are currently more infected men than women, thus increasing the probability that a woman would have an infected partner, especially in neighborhoods with high concentrations of infected men.

Although there is a pressing need to develop programs aimed at reducing HIV risk behaviors among women with injection drug–using partners, little is known about this group.[1,17] Even in the midst of the AIDS epidemic, drug treatment programs have made few efforts to work with partners of injection drug users.[18] If meaningful intervention programs are to be developed for women with injection drug using sex partners, more information is needed about their extent of risk and other aspects of their lives.

In this study, data from a large national cohort of women with injection drug–using sex partners were analyzed. As specified by the research protocol, the women reported one or more injection drug–using sex partners and no such drug use themselves in the 6 months prior to their initial interview.

Previous analyses of an earlier version of this data set (n = 3589) focused on racial/ethnic differences among the women in sexual risk, drug use, and incidence of sexually transmitted diseases.[8] Major racial/ethnic groups were African Americans, Whites, Puerto Rican Latinas, and other Latinas (primarily Mexicans). Results indicated that, for women with one injection drug–using partner, White women reported a higher frequency of unprotected vaginal sex than did those in the other groups. Among

The authors are with the National Development and Research Institutes, New York, NY, and the National AIDS Research Consortium.

Requests for reprints should be sent to Stephanie Torta, PhD, National Development and Research Institutes Inc, 11 Beach St, New York, NY 10013.

This paper was accepted November 24, 1993.

women with multiple partners, there were no racial/ethnic differences in mean frequency of unprotected vaginal sex. African Americans were more likely to report using crack, alcohol, and cocaine than were those in the other groups. African Americans and Whites generally reported more sexually transmitted diseases than did women in either Latina group.

Because patterns of sexual activity vary and may have a major impact on the degree of AIDS risk as well as other health and economic factors, it was decided to compare three behaviorally defined groups of heterosexual women with injection drug-using partners. These groups were composed of women who, during the 6 months prior to the interview, (1) had single partners, (2) had multiple partners and did not exchange sex for drugs and/or money, and (3) had multiple partners and did exchange sex for drugs and/or money. The three groups were compared on demographic and economic characteristics, involvement with the criminal justice system, sexual risk behaviors, non-injected drug use, and health characteristics. The effect of the grouping on HIV risk and other variables was then examined while controlling for selected demographic characteristics.

Methods

Procedure

The National AIDS Demonstration and Research program, on which this research was based, was implemented in 1987 by the National Institute on Drug Abuse and targeted two "hard to reach" groups: out-of-treatment injection drug users and those with injection drug-using sex partners. The goals were to collect information concerning HIV risk and to test the efficacy of various AIDS prevention strategies. Sixty-three sites were involved across the continental United States, in Hawaii and Puerto Rico, and along the US-Mexican border. Both target groups were offered equivalent incentives for participation. This paper reports only on heterosexual women with injection drug-using sex partners.

Individual sites determined their own recruitment strategies but were generally guided by indigenous outreach workers with knowledge of local drug using networks. To maximize the participation of this "hidden population," recruitment efforts were directed toward "street" locations where drug users congregate, as well as places where their partners are likely to be found (e.g., family-planning/obstetrical-gynecological clinics, housing projects, shelters, social service agencies, prisons). Efforts were made to recruit sex workers because many have injection drug-using partners.[7] Prior to any intervention, the women participated in a structured 45-minute interview. Topics included demographic and economic characteristics, involvement with the criminal justice system, sexual risk behaviors, non-injection drug use, incidence of sexually transmitted diseases, and other health characteristics. Data were collected from January 1988 through June 1991. Test-retest reliability (over an interval of 2 to 14 days; n = 196) was high for questions on number, gender, and injection drug-using status of sex partners and on exchange of sex for money or drugs (kappa range of .53 to 1.00). Test-retest correlations for the most risky sexual behaviors were .66 (vaginal sex) and .79 (anal sex). Test-retest correlations for frequency of noninjection drug use, based on ever having used a particular drug, ranged from .15 for heroin to .75 for alcohol. Women were offered HIV testing, and, when they were tested by National AIDS Demonstration and Research projects, HIV serostatus was included in the data set. In order to minimize responding based on "self-diagnosis," questions regarding history of sexually transmitted diseases were worded as follows: "Have you ever been told by a doctor or a nurse that you had . . .?" Systematic quality control procedures were implemented by a research firm under contract to the National Institute on Drug Abuse to verify client eligibility criteria and to assess the consistency of responses within each questionnaire.

Sample

The sample consisted of 5162 heterosexual women with injection drug-using partners; 53.1% were African Americans, 15% were Puerto Rican Latinas, 15% were Mexican Latinas, 12.5% were Whites, and 4.4% were Asian Americans, Native Americans, or other Latinas. The geographic distribution was as follows: 30% from the South, 27% from the Northeast, 23% from the West, 11% from Puerto Rico, and 10% from the North Central region.

Only 43% were high school graduates, and 15.5% were employed (full or part time). Homeless women (living on the streets or in a shelter) accounted for 11.6% of the sample. Regarding sexual activity, 53% (n = 2734) reported a single injection drug-using partner, 23.3% (n = 1205) reported multiple partners but had not exchanged sex for drugs and/or money, and 23.7% (n = 1223) reported multiple partners and had exchanged sex for drugs and/or money.

Statistical Analyses

One-way analyses of variance (ANOVAs) and chi-square tests were used to compare the three groups on demographics, drug use, sexual behavior, and health characteristics. Pairwise comparisons of means were conducted with Tukey's studentized range test when a significant ($P < .05$) ANOVA F value was obtained. All chi-square tests producing a significant overall finding ($P < .05$) were followed up with a series of 2×2 chi-squares to compare all three possible pairs of groups. Chi-square follow-up tests were reported as significant at the $P < .05$ level or better.

A second stage of analysis was conducted to assess whether grouping predicted risk behaviors and health outcomes when major demographic variables were controlled. Both multiple regression (sex risk and drug use behaviors) and logistic regression (dichotomous health and AIDS outcomes) were used. The grouping effect was examined in two ways: (1) by comparing women who had multiple injection drug-using partners but who did not exchange sex for drugs and/or money and those who had single partners and (2) by comparing women who had multiple injection drug-using partners and who did exchange sex for drugs and/or money and those who had multiple partners but did not exchange sex for drugs and/or money. Four series of regression analyses were conducted, since sex risk and drug use behaviors (continuous variables) and health and AIDS outcomes (dichotomous variables) were analyzed separately for each of the two group comparisons. The following variables were included in all regression analyses: age, African-American ethnicity, Latina ethnicity, northeastern region, southern region, western region, homelessness, and high school education. These independent variables (except for age) were coded dichotomously.

Results

Significance levels for the variables discussed in this section are reported only in the tables.

Demographic Characteristics

Table 1 indicates many significant differences among the groups. Women

with multiple partners who exchanged sex for drugs and/or money were most likely from the Northeast (35.8%); women with single partners were most likely from the South (32.8%) (χ^2 = 222.47). While women in all three groups were most likely to be African-American (χ^2 = 311.60), there were significant differences by group: two thirds of women with multiple partners were African-American, compared with about one half of the other groups. About 40% of women with single partners were Latina, in comparison with less than 20% of women who exchanged sex for drugs and/or money. Women with single partners (mean age = 30.3 years) were slightly older than those in either of the other two groups (F = 22.94, df = 25159) (see Table 1 for more detail).

Women with multiple partners who exchanged were more likely to be homeless (17.9%) than either women with single partners (7.5%) or those with multiple partners (11%) (χ^2 = 95.20). They were less likely to be high school graduates (40.0%) than either women with single partners (41.7%) or women with multiple partners (48.8%) (χ^2 = 22.98) and less likely to have children living with them (35.6%) than either women with single partners (66.5%) or women with multiple partners (60.1%) (χ^2 = 334.77). Women with multiple partners who exchanged were also more likely to have been in drug treatment (41.7%) than either women with single partners (15.3%) or those with multiple partners (24.7%) (χ^2 = 324.61).

TABLE 1—Demographic Characteristics of Women with Injection Drug–Using Sex Partners

	Single Sex Partner (Group 1) (n = 2734)	Multiple Partners without Exchange (Group 2) (n = 1205)	Multiple Partners with Exchange (Group 3) (n = 1223)	Group Differences
Geographic region,** %				
Northeast	23.7	25.2	35.8	...
North Central	9.2	12.7	7.9	...
South	32.8	26.5	26.3	...
West	19.4	25.1	27.3	...
Puerto Rico	14.9	10.6	2.8	...
Ethnic status,** %				
African American	47.0	50.8	69.0	...
Puerto Rican	20.0	13.0	5.8	...
Mexican American	18.2	11.0	11.9	...
White	10.2	19.4	10.9	...
Other	4.6	5.8	2.4	...
Mean age, y (SD)*	30.3 (8.3)	28.9 (7.7)	28.6 (6.7)	1 > 2, 3
Homeless, %	7.5	11.0	17.9	3 > 2 > 1
High school graduates, %	41.7	48.8	40.0	2 > 1, 3

Note. All reported group differences were significant at $P < .01$ or better.
*$P < .001$; **$P < .0001$.

TABLE 2—Economic Characteristics and Involvement with the Criminal Justice System: Women with Injection Drug–Using Sex Partners

	Single Sex Partner (Group 1) (n = 2734), %	Multiple Partners without Exchange (Group 2) (n = 1205), %	Multiple Partners with Exchange (Group 3) (n = 1223), %	Group Differences
Full- or part-time job**	16.9	16.7	11.6	1, 2 > 3
Source of income during previous 6 mo[a]				
Illegal activities**	5.8	8.6	57.8	3 > 2 > 1
Welfare**	67.5	66.1	55.4	1, 2 > 3
Sex partner(s)**	28.2	25.3	44.0	3 > 1, 2
Friend(s)**	9.2	19.7	28.7	3 > 2 > 1
Job**	29.0	35.2	27.4	2 > 1, 3
Family member(s)**	20.7	26.1	26.2	2, 3 > 1
Other sources*	30.9	35.5	34.8	2, 3 > 1
Ever been in jail or prison**	32.0	38.1	58.8	3 > 2 > 1
Legal status				
On probation**	6.9	9.0	14.8	3 > 2 > 1
On parole	1.3	1.8	2.2	
Facing charges**	3.5	3.8	10.6	3 > 1, 2

Note. All reported group differences were significant at $P < .01$ or better.
[a]Response categories are not mutually exclusive.
*$P < .001$; **$P < .0001$.

Economic Characteristics and Criminal Justice Involvement

Table 2 displays information regarding economic characteristics and experiences with the criminal justice system. Both women with single partners (16.9%) and those with multiple partners (16.7%) were more likely to be employed than were women with multiple partners who exchanged sex for drugs and/or money (11.6%) (χ^2 = 19.87). Women with multiple partners who exchanged (57.8%) were more likely to report income from illegal sources than were either women with single partners (5.8%) or those with multiple partners (8.6%) (χ^2 = 1582.69). Although most women with multiple partners who exchanged (55.4%) reported income from welfare, the proportion was still significantly less than the proportions of women with single partners (67.5%) and women with multiple partners (66.1%) (χ^2 = 51.68). More women with multiple partners who exchanged reported income from sex partners (44%) than either women with single partners (28.2%) or those with multiple partners (25.3%) (χ^2 = 124.10).

Women with multiple partners who exchanged sex for drugs and/or money were also most likely to have had experience with the criminal justice system. More than 58% of these women had been

TABLE 3—Sexual Behavior of Women with Injection Drug–Using Partners

	Single Sex Partner (Group 1) (n = 2734)	Multiple Partners without Exchange (Group 2) (n = 1205)	Multiple Partners with Exchange (Group 3) (n = 1223)	Group Differences
Frequency of vaginal sex without a condom[a]**	22.7	24.2	34.5	3 > 1, 2
Frequency of oral sex without a condom[a]**	4.6	7.7	19.5	3 > 2 > 1
Frequency of anal sex without a condom[a]*	1.2	1.3	2.4	3 > 1, 2
Total no. unprotected sexual acts[b]**	28.6	33.1	56.5	3 > 2 > 1
Condom use,[c]** %	7.8	15.5	34.0	3 > 2 > 1
Mean no. drug injecting partners	...	1.4	9.5	3 > 2

Note. All reported group differences were significant at $P < .01$ or better.
[a]Mean monthly frequency during the 6 months prior to the initial interview.
[b]Sum of vaginal, oral, and anal sex (without a condom) during the 6 months prior to the interview.
[c]Percentage of time condoms were used during vaginal, oral, and anal sex in the 6 months prior to the interview.
*$P < .001$; **$P < .0001$.

TABLE 4—Drug Use Behavior of Women with Injection Drug–Using Partners

	Single Sex Partner (Group 1) (n = 2734)	Multiple Partners without Exchange (Group 2) (n = 1205)	Multiple Partners with Exchange (Group 3) (n = 1223)	Group Differences
Frequency of crack use**	7.3	10.0	50.1	3 > 1, 2
Frequency of alcohol use**	12.9	18.2	31.6	3 > 2 > 1
Frequency of noninjected cocaine use**	4.5	6.1	22.4	3 > 1, 2
Frequency of marijuana use**	8.0	12.2	17.5	3 > 2 > 1
Frequency of noninjected heroin use*	2.1	2.3	3.7	3 > 1, 2

Note. Results are reported in terms of mean monthly frequencies during the 6 months prior to the interview. All reported group differences were significant at $P < .01$ or better.
*$P < .05$; **$P < .0001$.

imprisoned, in comparison with 32% of women with single partners and 38.1% of women with multiple partners (χ^2 = 254.80). Women with multiple partners who exchanged (10.6%) were also more likely to be facing criminal charges at the time of the interview than were women with multiple partners (3.8%) or women with single partners (3.5%) (χ^2 = 90.38).

Sexual Risk Behaviors

Three sexual behaviors were examined (see Table 3): frequency of unprotected vaginal sex, oral sex (fellatio), and anal sex. Also included was the sum of the three unprotected sexual activities and the mean percentage of condom use for the three activities. Although women with multiple partners who exchanged sex for drugs and/or money (34%) reported a higher mean monthly percentage of condom use across all three activities than did women with single partners (7.8%) or those with multiple partners (15.5%) (F = 486.80, df = 2512), the frequency of unprotected sex in this group was also significantly higher than in the other groups. Mean monthly frequency of unprotected vaginal sex among women with multiple partners who exchanged was 34.5, significantly more than women with single partners and women with multiple partners (22.7 and 24.2, respectively; F = 548.11, df = 2515). Unprotected anal sex was a relatively low frequency behavior. Women with multiple partners who exchanged (56.5) reported a higher total number of unprotected sex acts than did women with single partners (28.6) and women with multiple partners (33.1) (F = 123.81, df = 2515). Finally, women with multiple partners who exchanged had significantly more injection drug–using partners (mean = 9.5) than did women with multiple partners (mean = 1.4).

Noninjection Drug Use

During the 6 months prior to the interview, the mean monthly frequencies of crack, alcohol, cocaine, marijuana, and noninjection heroin use, analyzed individually for each substance, were highest for women with multiple partners who exchanged sex for drugs and/or money. The most frequently used drugs were crack, alcohol, and noninjected cocaine. As Table 4 indicates, the frequency of crack use for women with multiple partners who exchanged was extremely high (50.1); among women with single partners and women with multiple partners, the frequencies were considerably lower (7.3 and 10, respectively). The mean monthly frequency of alcohol use among women with multiple partners who exchanged was 31.6; the corresponding figures for women with single partners and those with multiple partners were 12.9 and 18.2 (F = 110.97, df = 2515). Women with multiple partners who exchanged (22.4) also reported more frequent noninjected cocaine use than did either women with single partners (4.5) or those with multiple partners (6.1) (F = 170.5, df = 2,515).

Health Characteristics

The health characteristics assessed (see Table 5) were prevalence of gonorrhea, syphilis, chlamydia, genital herpes, genital sores, and persistent yeast infections; perceived likelihood of contracting AIDS; perceived health status; pregnancy status; and HIV serostatus. The most prevalent sexually transmitted disease was gonorrhea; 40.1% of women with multiple partners who exchanged reported ever having gonorrhea, in comparison with 25.8% of women with multiple partners and 13.4% of women with single partners (χ^2 = 369.32). Women with multiple partners who exchanged were also more likely to report ever having syphilis (17.7%) than women with multiple partners (7.9%) or women with single partners (4.1%) (χ^2 = 206.69). Finally, women with mul-

tiple partners who exchanged (22.5%) were more likely than women with multiple partners (13%) and those with single partners (8.1%) to report ever having genital sores ($\chi^2 = 159.45$).

More women with multiple partners who exchanged reported being pregnant (16.43%) than either women with single partners (11.98%) or those with multiple partners (9.93%) ($\chi^2 = 23.71$). Approximately 30% of the sample underwent HIV testing. Women with single partners and women with multiple partners (both 33.7%) were more likely to report having had an HIV test than women with multiple sex partners who exchanged (20.8%) ($\chi^2 = 73.58$). Although more women with multiple partners who exchanged (7.9%) were confirmed as seropositives than women with single partners (5.4%) and women with multiple partners (4.4%), the differences were not statistically significant ($\chi^2 = 3.60$). Finally, among women who were not tested, those with multiple partners who exchanged (21.4%) were more likely to report that they had a high or sure chance of contracting AIDS than were either women with single partners (13%) or women with multiple partners (11%) ($\chi^2 = 41.73$).

Effects of Group Status

As Table 6 indicates, most of the differences in sex risk and drug use behaviors between women with multiple sex partners who exchanged sex for drugs and/or money and women with multiple partners who did not do so remained highly significant, even when the effects of demographic variables were taken into account. Thus, the effect of exchanging was significant above and beyond the combined influence of age, ethnicity, geographic region, level of education, and homelessness. However, the effect of having multiple sex partners (vs having one partner) was less pronounced. This effect remained significant for alcohol and marijuana use, but there were no differences between women with multiple partners and women with single partners in the use of crack, cocaine, or heroin or in unprotected sex when demographic characteristics were controlled.

The results presented in Table 7 suggest that the effects of (1) having multiple sex partners vs a single sex partner and (2) having multiple partners and exchanging sex for drugs and/or money (among women with multiple partners) are both associated with a significantly greater likelihood of having a sexually transmitted disease and having had each of several specific illnesses. Thus, women with multiple sex partners were about twice as likely (odds ratio [OR] = 2.16) as women with single sex partners to have had at least one sexually transmitted disease, and, among women with multiple partners, those who exchanged (OR = 1.80) were more likely than those who did not to have had a sexually transmitted disease, even when demographic factors were taken into account. The effects of having multiple partners vs having a single partner and of exchanging vs not exchanging were also associated with a significantly greater likelihood of having had genital herpes, gonorrhea, syphilis, and genital sores.

TABLE 5—Health Characteristics of Women with Injection Drug–Using Sex Partners

	Single Sex Partner (Group 1) (n = 2734), %	Multiple Partners without Exchange (Group 2) (n = 1205), %	Multiple Partners with Exchange (Group 3) (n = 1223), %	Group Differences
Prevalence of sexually transmitted diseases				
Gonorrhea (ever)*	13.4	25.8	40.1	3 > 2 > 1
Syphilis (ever)*	4.1	7.9	17.7	3 > 2 > 1
Chlamydia (ever)*	4.6	10.9	12.7	2, 3 > 1
Genital herpes (ever)*	1.5	3.4	6.3	3 > 2 > 1
Any sexually transmitted disease (ever)*	19.8	36.9	53.5	3 > 2 > 1
Prevalence other conditions				
Genital sores (ever)*	8.1	13.0	22.5	3 > 2 > 1
Persistent yeast infection (ever)*	6.2	6.5	12.0	3 > 1, 2
Perceived health status*				
Fair or poor	36.1	39.1	47.8	3 > 2 > 1
Good or excellent	63.9	60.9	52.2	
Reported being pregnant*	12.0	9.9	16.4	3 > 1, 2
HIV testing				
Tested	33.7	33.7	20.8	1, 2 > 3
Tested positive[a]	5.4	4.4	7.9	
Perceived likelihood of contracting AIDS[b]*: high or sure chance	13.0	11.0	21.4	

Note. All reported group differences were significant at $P < .01$ or better.
[a]Results are based on the 1582 persons (922 in group 1, 406 in group 2, and 254 in group 3) who had one HIV test under the auspices of the National AIDS Demonstration and Research programs.
[b]Results are based on the 3580 persons (1812 in group 1, 799 in group 2, and 969 in group 3) who were not tested as part of the National AIDS Demonstration and Research programs.
*$P < .0001$.

Discussion

Several points must be made before considering the implications of these data. First, National AIDS Demonstration and Research projects were not planned as population studies, and subjects were not randomly selected. Thus, the data were not intended to represent the universe of women with injection drug–using sex partners. Also, the data, excluding HIV serostatus, were based on self-report. There may be limitations to the precision of these data, especially since most questions concerned behavior over a period of 6 months. Finally, in order to have been involved in these projects, women had to have some knowledge of a partner's injection drug use. As noted previously, many women probably are unaware of their partner's injection drug use and thus constitute one particularly vulnerable group not represented here. Nonetheless, conclusions drawn from this research can be helpful in planning prevention strategies as well as indicating areas for future research.

TABLE 6—The Effect of Sex Partner Status on Sex Risk and Drug Use Behaviors, with Age, Ethnic Status, Geographic Region, Education, and Homelessness Controlled

Risk Behavior	Multiple Partners without Exchange vs Single Sex Partner		Multiple Partners with Exchange vs Multiple without Exchange	
	Parameter Estimate[a]	*t*	Parameter Estimate[b]	*t*
Unprotected sex acts (vaginal, oral, and anal), no.	2.93	1.92	23.20	8.82***
Condom use, %	6.74	8.86***	16.91	14.06***
Alcohol use	4.33	3.87**	12.47	7.08***
Marijuana use	2.90	3.11*	4.73	3.31**
Crack use	1.85	1.91	33.51	16.94***
Noninjection cocaine use	1.09	1.50	13.71	8.92***
Noninjection heroin use	0.48	0.91	1.28	1.76

Note. All risk behaviors (except condom use) are expressed as mean number of times per month.
[a]Statistically adjusted mean difference between women with multiple partners (no exchange) (n = 1205) and women with single sex partners (n = 2734).
[b]Statistically adjusted mean difference between women with multiple partners who exchanged (n = 1223) and women with multiple partners (no exchange) (n = 1205).
*$P < .05$; **$P < .001$; ***$P < .0001$.

TABLE 7—The Effect of Sex Partner Statuš on Health and AIDS-Related Outcomes, with Age, Ethnic Status, Geographic Region, Education, and Homelessness Controlled

Outcome	Multiple Partners without Exchange vs Single Sex Partner		Multiple Partners with Exchange vs Multiple Partners without Exchange	
	Odds Ratio[a]	Wald Chi-Square	Odds Ratio[b]	Wald Chi-Square
Genital herpes (ever)	2.03	9.38**	2.16	13.43***
Gonorrhea (ever)	2.16	71.88****	1.71	33.65***
Syphilis (ever)	1.98	21.27***	2.07	29.08***
Chlamydia (ever)	2.00	26.18***	1.28	3.29
Any sexually transmitted disease (ever)	2.16	92.95****	1.80	45.07***
Genital sores (ever)	1.53	13.84***	1.96	34.13***
Yeast infection (ever)	1.00	0.00	2.11	22.94***
Perceived health status of fair or poor	1.16	4.02*	1.46	19.36***
Pregnant at time of interview	0.67	10.28**	1.45	7.16**
Tested one or more times for HIV	0.95	0.42	0.51	43.56***
Tested positive on at least one HIV test	0.88	0.17	1.37	0.78
High or sure perceived likelihood of contracting AIDS	0.89	1.40	1.92	41.21***

Note. All outcome variables were coded dichotomously.
[a]Statistically adjusted difference in the likelihood of having a particular outcome between women with multiple partners (no exchange) (n = 1205) and women with single partners (n = 2734).
[b]Statistically adjusted difference in the likelihood of having a particular outcome between women with multiple partners who exchanged (n = 1223) and women with multiple partners (no exchange) (n = 1205).
*$P < .05$; **$P < .01$; ***$P < .001$; ****$P < .0001$.

Overall, considerable numbers of these women were poorly educated and unemployed. Also, many had jail or prison experiences, were homeless, and had significant drug problems. These facts suggest that their daily need to survive may take precedence over other motivations, including the prevention of AIDS. AIDS is but one of many concerns faced by these women, most of whom reside in economically and socially devastated inner cities. Experience with the National AIDS Demonstration and Research projects indicates that women with injection drug–using partners can be reached and appear interested in reducing their HIV risk.[19] However, it must be recognized that the more immediate goal of HIV prevention may not be realized without addressing the broader social issues affecting these women's lives.

The major conclusion reached from these data is that differences among women with injection drug–using partners are clearly associated with the type of sexual contacts they have with men. On nearly all of the characteristics examined, even when major demographic variables were controlled, findings indicated that women who traded sex for drugs and/or money lead lives that are the most chaotic, the least healthy, and the most risky. The differences are not only statistically significant but are of a magnitude that is very dramatic in terms of what they mean in "real life." As poignantly written by researchers involved in a Miami-based project, "women who exchange sex for crack [as do many in this group] experience a level of human suffering previously unknown in the street drug scene."[20] Women who exchanged were much more likely to be homeless, to have been incarcerated, and to have engaged in illegal activities as a way to survive. Although they were more likely to report that their partners used a condom during sexual activities, a better indication of HIV risk was the significantly higher amount of unprotected high risk (vaginal and anal) sex in which they engaged. Also, they had significantly more injection drug–using partners than did women in the other groups, further increasing their risk for exposure to the virus. Crack use was rampant in this group, indicating a pressing need for drug treatment as a part of HIV prevention. Also, the rate of sexually transmitted diseases was extremely high, probably as a result of urban poverty, social disintegration, prostitution, and the exchange of sex for drugs.[21] The rate of seropositivity in this group was higher than that in the other two groups, but not significantly so. However, because serostatus results were available for only 30% of the sample, it is highly likely, given the levels of risk reported among this group, that the difference would have been

significant if a larger number of women had been tested.

Much that has been written about sex workers and AIDS has focused on the role of the former as "vectors of transmission" into the heterosexual community.[1] The dangers of this type of life for the women themselves have often not been addressed. Yet, with the relative efficiency of transmission from an infected man to a woman and the high probability in certain communities that women have sexual contacts with large numbers of infected men, these women probably face a greater risk of being infected themselves than infecting others.

Women with single partners presented a different picture in terms of HIV risk. They were most likely to receive an income from public assistance and to have children living with them. They were least likely to have been in jail or prison or in a drug treatment program. Their men were also least likely to use a condom during sexual activities. Thus, although they had a lower frequency of unprotected vaginal sex than the women with multiple partners who exchanged sex for drugs and/or money, women with a single injection drug–using sex partner must still be considered at high risk. Because their reports of noninjected drug use were relatively low, as was their incidence of sexually transmitted diseases, they are mainly at risk through unprotected sex with a single injection drug user. While condom use is recommended as an HIV prevention strategy for all sex partners of injection drug users, qualitatively different relationships (e.g., monogamous couples vs women who trade and their clients) may call for different approaches to condom use negotiation. This issue needs more investigation. Also recommended is more research on developing and testing protective methods that are controlled by women, including the recently introduced female condom.[22]

Women with multiple sex partners who did not trade sex for drugs and/or money appeared to be more like women with single sex partners on both demographic characteristics and level of HIV risk. Thus, they are at risk primarily through unprotected sex with multiple injection drug users.

Many research questions remain to be investigated. For example, which demographic and drug use variables predict group status? Also, more information is needed regarding the relative efficacy of different models of outreach and intervention with these women. HIV risk reduction efforts among drug users have generally focused on two behavior change strategies: social learning techniques (to teach the skills necessary for risk reduction) and the community organizing model, which is based on changing norms, values, and practices directly connected with HIV risk reduction.[23] It is not yet clear which is most relevant to women with injection drug using sex partners. Finally, the present study focused on a single point in time. One particular variable that needs to be examined more closely is the past use of injection drugs by this population. Unfortunately, detailed information regarding prior injection behavior, including when drug injection stopped, was not available from this sample. Additional research is required to assess the extent to which risk behaviors among identified groups of women with injection drug–using partners vary over time and to identify those factors affecting change. □

Acknowledgments

This research was supported by the National Institute on Drug Abuse through the program of National AIDS Demonstration and Research grants. The research was done in collaboration with the principal investigators of the National AIDS Research Consortium. (For a list of the investigators, please contact the first author.)

We gratefully acknowledge the assistance of NOVA Research Company as the national data coordination center. We also thank Dr Barry Brown, for his leadership in coordinating this national effort, and Ms Gloria Weissman, for her commitment to HIV prevention for women.

References

1. Hankins CA, Handley MA. HIV disease and AIDS in women: current knowledge and a research agenda. *J AIDS.* 1992;5:957–971.
2. Cohen JB. Why woman partners of drug users will continue to be at high risk for HIV infection. *J Addict Dis.* 1991;10:99–110.
3. Miller HG, Turner CF, Moses LC. *AIDS: The Second Decade.* Washington, DC: National Academy Press; 1990.
4. *HIV/AIDS Surveillance.* Atlanta, Ga: Centers for Disease Control; 1993.
5. Cohen J, Hauer L, Wofsy C. Women and IV drugs: parenteral and heterosexual transmission of human immunodeficiency virus. *J Drug Issues.* 1988;19:39–56.
6. Chitwood DD, Inciardi JA, McBride DC, McCoy CB, McCoy HV, Trapido E. *A Community Approach to AIDS Intervention.* Westport, Conn: Greenwood Press; 1991.
7. Corby N, Wolitski R, Thornton-Johnson S, Tanner W. AIDS knowledge, perception of risk, and behaviors among female sex partners of injection drug users. *AIDS Educ Prev.* 1991;3:353–366.
8. Deren S, Davis WR, Tortu S, Ahluwahlia I. Characteristics of female sexual partners. In: Brown B, Beschner G, National AIDS Research Consortium, eds. *At Risk for AIDS: Drug Injection Users and Their Sexual Partners.* Westport, Conn: Greenwood Press. In press.
9. Chiasson MA, Stoneburner R, Hildebrandt D, Ewing W, Telzak E, Jaffe H. Heterosexual transmission of HIV-1 associated with the use of smokable freebase cocaine (crack). *AIDS.* 1991;5:1121–1126.
10. Booth RE, Watters JK, Chitwood DD. HIV risk–related sex behaviors among injection drug users, crack smokers, and injection drug users who smoke crack. *Am J Public Health.* 1993;83:1144–1148.
11. Sterk C. Cocaine and HIV seropositivity. *Lancet.* 1988;1:1052–1053. Letter.
12. Minkoff HL, McCalla S, Delke I, et al. The relationship of cocaine use to syphilis and human immunodificiency virus infection among inner city parturient women. *Am J Obstet Gynecol.* 1990;163:521–526.
13. Greenberg MSZ, Singh T, Htoo M, Shultz S. The association between congenital syphilis and cocaine/crack use in New York City: a case-control study. *Am J Public Health.* 1991;81:1316–1318.
14. Marx R, Aral SO, Rolfs RY, et al. Crack, sex, and STD. *Sex Transm Dis.* 1991;18:92–101.
15. Cates W, Hinman AR. Sexually transmitted diseases in the 1990's. *N Engl J Med.* 1991;325:1368–1370.
16. Padian NS, Shiboski SS, Jewell N. The relative efficiency of female-to-male HIV sexual transmission. Presented at the Sixth International Conference on AIDS, June 1990, San Francisco, Calif.
17. Weissman G, National AIDS Research Consortium. AIDS prevention for women at risk: experience from a national demonstration research project. *J Primary Prev.* 1991;12:49–63.
18. Brown BS, Beschner GM. AIDS and HIV infection—implications for drug abuse treatment. *J Drug Issues.* 1989;19:149–150.
19. Stephens RC, Simpson DD, Coyle SL, McCoy CB, National AIDS Research Consortium. Comparative effectiveness of NADR interventions. In: Brown BS, Beschner GM, National AIDS Research Consortium, eds. *Handbook on Risk of AIDS: Injection Drug Users and Their Sexual Partners.* Westport, Conn: Greenwood Press. 1993.
20. Inciardi JA, Lockwood D, Pottieger AE. Crack-dependent women and sexuality: implications for STD acquisition and transmission. *Addict Recovery.* 1991;(July/August):25–28.
21. Aral SO, Holmes KK. Sexually transmitted diseases in the AIDS era. *Sci Am.* 1991;264:62–69.
22. Gollub EL, Stein ZA. Commentary: the new female condom—item 1 on a women's AIDS prevention agenda. *Am J Public Health.* 1993;83:498–500.
23. Friedman SR, Neaigus A, Des Jarlais DC, et al. Social intervention against AIDS among injecting drug users. *Br J Addict.* 1992;87:393–404.

A Recent Decline in Cocaine Use among Youthful Arrestees in Manhattan, 1987 through 1993

ABSTRACT

Objectives. Cocaine use among youths as measured by several annual surveys was most popular from about 1979 to 1986, after which it declined. This study carefully examines the nature of the decline by focusing on microdata for youthful arrestees in Manhattan.

Methods. Multiple statistical analyses examine whether the decline in cocaine use detected by urinalysis is attributable to fewer arrested youths born more recently having become regular users (cohort effect), to regular users decreasing their consumption (period effect), or to changes in arrest or sampling priorities (artifact).

Results. All analyses suggest that the dramatic decline in detected cocaine use among arrestees—from 69% in 1987 to 17% in 1993—was a cohort effect. Detected cocaine use, which was highest (78%) among arrestees reaching 18 in 1986 at the height of the crack epidemic in New York City, subsequently declined to a low of 10% among arrestees reaching 18 in 1993.

Discussion. These findings suggests that the epidemic in use of cocaine and crack entered a decline in the late 1980s. However, widespread use of these drugs will probably continue to prevail as an aging population with established habits persists in its use. (*Am J Public Health.* 1994;84:1250–1254)

Andrew Golub, PhD, and Bruce D. Johnson, PhD

Introduction

This paper examines recent trends in illicit drug use among very high-risk youths who have problems with both drug use and the law. Persons tend to establish lifetime drug use patterns by their teens or mid-20s.[1–4] Many adults eventually stop using certain drugs or decrease their frequency of use. However, some persons who begin using or abusing illicit drugs as teens persist in such use for a long time. Hence, trends in drug use among youths provide a harbinger of future drug use patterns by these persons as adults. For example, in New York City today, the majority of heroin injectors are currently in their late 30s, 40s, and 50s, and began injecting heroin in the 1960s and early 1970s.[5–10]

Table 1 reports trends in the use of certain illegal drugs among youths from three major ongoing surveys. Monitoring the Future obtains self-reported drug use histories from a national probability sample of high school seniors.[11] The National Household Survey on Drug Abuse uses a nationwide sample of households.[12,13] The Drug Use Forecasting (DUF) program, started in 1987, collects urine specimens along with self-reported information from a sample of arrestees at each of 24 central booking facilities nationwide.[14–16] Results from the first two surveys exhibit comparable trends in drug use over time; those from the Drug Use Forecasting program in Manhattan reveal much higher levels of drug use than the other two surveys.

A recent report by the US General Accounting Office suggests that the overall prevalence of drug use measured by each of these three surveys can be misleading.[17] Self-reported data from Monitoring the Future and the National Household Survey may be biased downward by subjects' unwillingness to report drug use and by the fact that many serious drug abusers neither remain in school nor live in stable households. The Drug Use Forecasting program obtains an objective measurement of recent drug use with urine samples. Through use of the enzyme-multiplied immunoassay technique urinalysis screen, it detects cocaine use within the past 48 to 72 hours but does not distinguish mode of use (i.e., snorting, smoking, or injecting).[15] Consequently, many youthful arrestees detected as recent cocaine users may have been crack users.

The Drug Use Forecasting sample of arrestees, however, is not representative of the general community, and so the prevalence of detected drug use within it is typically much higher.[17] Most of its youthful arrestees in Manhattan were charged with a serious crime, 80% for a felony; additionally, most are male (82%) and minority (80% Black or Hispanic).

This study focuses less on the overall magnitude of drug use and more on changes over time, especially the decline in the detected use of cocaine from 69% in 1987 to 17% in early 1993 among Drug

Andrew Golub is with the Department of Public Management, John Jay College of Criminal Justice, The City University of New York, New York, NY. In addition, Dr Golub and Bruce D. Johnson are with National Development and Research Institutes, Inc, New York, NY.

Requests for reprints should be sent to Andrew Golub, PhD, John Jay College, Department of Public Management, 445 W 59th St, New York, NY 10019.

This paper was accepted April 12, 1994.

Note. The views expressed here are the authors' and do not necessarily represent the official positions of the US government, National Development and Research Institutes, Inc, or John Jay College of Criminal Justice.

TABLE 1—Trends in Youthful Drug Use as Reported by Drug Use Forecasting–Manhattan, Monitoring the Future, and the National Household Survey on Drug Abuse, 1975 through 1993

	Percentage Reporting Use, by Interview Year																		
	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
Drug Use Forecasting[a]																			
Cocaine/crack													69	61	51	31	25	23	17[b]
Opiates													14	8	8	5	2	8	4[b]
Marijuana/hashish													35	34	27	28	31	40	24[b]
Any of 10 drugs[c]														71	62	51	50	56	39[b]
Monitoring the Future[d]																			
Cocaine/crack	1.9	2.0	2.9	3.9	5.7	5.2	5.8	5.0	4.9	5.8	6.7	6.2	4.3	3.4	2.8	1.9	1.4	1.3	1.3
Heroin	0.4	0.2	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.2	0.2	0.2	0.3	0.2	0.2	0.3	0.2
Marijuana/hashish	27.1	32.2	35.4	37.1	36.5	33.7	31.6	28.5	27.0	25.2	25.7	23.4	21.0	18.0	16.7	14.0	13.8	11.9	15.5
Any illicit drug	30.7	34.2	37.6	38.9	38.9	37.2	36.9	32.5	30.5	29.2	29.7	27.1	24.7	21.3	19.7	17.2	16.4	14.4	18.3
National Household Survey on Drug Abuse[e]																			
Cocaine/crack		2.0	3.7		9.3			6.8			7.6			4.5			2.2	2.0	1.8	
Heroin		0.0	0.0		0.0			0.0			0.0			0.0			0.1	0.1	0.2	
Marijuana/hashish		25.0	27.4		35.4			27.4			21.8			15.5			12.7	13.0	11.0	
Any illicit drug[f]					37.1			30.4			25.7			17.8			14.9	15.4	13.0	

[a]The Drug Use Forecasting–Manhattan program, which detects recent use via urinalysis, recruits around 200 arrestees under age 21 per year resulting in standard errors for drug use prevalence on the order of 3.0%.
[b]First quarter results only.
[c]Since 1988, the Drug Use Forecasting program has tested for cocaine, opiates, cannabis, amphetamines, barbiturates, benzodiazepines, methadone, methaqualone, propoxyphene, and phencyclidine.
[d]From Johnston et al.[11] Monitoring the Future interviews about 16 000 high school seniors each year regarding drug use within the past 30 days and thus obtains extremely accurate prevalence rates with standard errors on the order of 0.2%.
[e]From the National Household Survey on Drug Abuse.[12,13] The National Household Survey on Drug Abuse, which also is concerned with drug use within the past 30 days, typically surveys more than 1000 persons aged 18 to 25 in each sampling year, which results in standard errors on the order of 0.7%.
[f]Question initiated in the 1979 survey.

Use Forecasting Manhattan arrestees under age 21. The overall trends in the use of any illicit drugs, as shown in Table 1, primarily reflect changes in the popularity of marijuana—the most commonly used illicit drug—which peaked in 1979 and declined thereafter. A previous study reported in this Journal documented that these national trends are mirrored by those among high school students in New York State.[18]

Table 1 indicates that cocaine use among youths nationally also peaked by 1979. Unlike marijuana, however, cocaine maintained its popularity through about 1986, after which time its use declined precipitously. This is documented in other studies,[23] in which cocaine use increased from about 1979 and then declined starting in the late 1980s. Overall drug use and the use of heroin declined in the late 1980s, suggesting that these former cocaine users did not replace cocaine with other known hard drugs. The congruence of findings from these three extensive data sets, along with observations by ethnographers,[19–22] thus provides strong evidence of a *passing cocaine fad among youths* as well as among adults.

Methods

The observed decline in detected cocaine use among Drug Use Forecasting–Manhattan youthful arrestees from 1987 to the first quarter of 1993 (1Q93) may have resulted from either of three phenomena: (1) a decline in use, as youths who previously used cocaine cut back on their frequency of use or terminated use completely; (2) a decline in initiation, as fewer of those youths who were born more recently initiated regular cocaine use; or (3) a change in sample characteristics, as changes in either policing priorities or Drug Use Forecasting–Manhattan recruitment procedures resulted in there being fewer cocaine-using offenders recruited in more recent samples.

Three statistical techniques were used to analyze this decline: an age-period-cohort analysis, logistic regression, and postdiction. The age-period-cohort analysis examined a two-way table displaying trends in prevalence of detected cocaine use as a function both of birth year down the rows (cohort effect) and of interview year across the columns (period effect). Any detectable period effect reflected both changes in drug use over time and changes in Drug Use Forecasting sampling procedures. An age effect would be observed as any differences across diagonals.

Logistic regression differs from ordinary least-squares regression primarily in that the dependent variable represents the odds for a particular trait—in this case, detected cocaine use. (For a thorough discussion of the properties of logistic regression, see Hosmer and Lemeshow.[24]) Just like ordinary least-squares regression, logistic regression has the important property of estimating the variation associated with each independent variable simultaneously and, thereby, controlling for the influence of all other variables included in the analysis. To control for any changes in the sample composition over time, the following arrest and demographic variables were included in this analysis: (1) most serious arrest charge, (2) whether the arrest was for a misdemeanor or a felony, (3) sex, and (4) race/ethnicity. Birth year was included to measure changes across birth cohorts. Similarly, interview year was included to identify any residual decline

TABLE 2—Variation in Detected Cocaine Use, by Birth Year and Interview Year

Birth Year	Proportion Detected as Cocaine Users (SE), and Count, by Drug Use Forecasting Interview Year[a]							
	1987	1988	1989	1990	1991	1992	1993	Average[b]
1966		75 (6)	79 (6)	59 (7)	58 (6)	72 (7)	68 (11)	
	0	61	56	49	64	39	19	0
1967	62 (8)	***93 (7)***	70 (5)	63 (7)	78 (6)	61 (7)	71 (9)	71 (7)
	34	***14***	74	56	50	46	24	48
1968	88 (6)	67 (7)	***87 (9)***	53 (6)	57 (7)	79 (7)	68 (10)	78 (4)
	34	43	***15***	62	54	38	25	92
1969	64 (10)	63 (8)	57 (7)	***62 (4)***	65 (6)	58 (7)	50 (13)	60 (4)
	22	38	56	***13***	57	48	16	129
1970	72 (11)	64 (9)	60 (6)	47 (8)		47 (9)	45 (16)	58 (4)
	18	33	60	45	***0***	30	11	156
1971	40 (16)	46 (10)	45 (9)	37 (7)	49 (8)	***58 (15)***	15 (10)	45 (4)
	10	24	31	46	45	***12***	13	168
1972		28 (11)	29 (7)	18 (6)	20 (7)	27 (9)	***31 (13)***	24 (3)
		18	38	44	35	26	***13***	174
1973			32 (11)	30 (7)	20 (6)	20 (7)	44 (18)	26 (4)
			19	44	40	35	9	147
1974			0 ...[c]	10 (6)	22 (7)	17 (6)	0 ...[c]	16 (3)
			1	30	37	36	5	109
1975				0 ...[c]	11 (6)	16 (7)	0 ...[c]	10 (4)
				3	27	25	12	67
1976			0 ...[c]	33 (33)	0 ...[c]	13 (13)	0 ...[c]	10 (7)
			1	3	5	8	3	20
1977					0 ...[c]	100 ...[c]	0 ...[c]	20 (20)
					1	1	3	5
1978							0 ...[c]	0 ...[c]
							1	1
Average[b]	69 (4)	61 (4)	51 (3)	31 (3)	25 (3)	23 (4)	17 (6)	41 (1)
	118	170	221	228	190	143	46	1116

[a]The bold italic figures indicate when a birth cohort reached age 21.
[b]Averages include only those arrestees under age 21.
[c]Standard error is meaningless when prevalence is either 0% or 100%.

resulting from former users decreasing or terminating use, controlling for both changes in sample composition and variation across birth cohorts. Quarter was included to detect any seasonal variation.

Postdiction examined the extent to which differences across birth cohorts and sample composition alone and in combination could account for the observed decline. The first step was to estimate a logistic regression model with detected cocaine use as the dependent variable. The independent variables included the sample composition variables and/or birth year. This regression model explicitly *excluded* the Drug Use Forecasting interview year and quarter as independent variables in order to determine how variation over time in the sample composition and birth cohort variables could reproduce (or postdict) the observed decline in detected cocaine use. The second step was to determine each subject's probability of detected cocaine use by employing the coefficient estimates from the logistic regression model. The last step was to calculate the postdicted rate of detected cocaine use for each year as the sum of the appropriate individual, postdicted probabilities.

Results

The substantial variation in detected cocaine use across birth years and the lack of a decline across interview years by birth cohort reported in Table 2 suggest that the decline in detected cocaine use was predominately a cohort effect and that few cocaine-using criminal offenders reduced or terminated use of cocaine. Overall, detected cocaine use declined dramatically across birth cohorts from a high of 78% among arrestees born in 1968 to a low of 10% among those born in 1975 and 1976.

The logistic regression analysis further documents the contribution of each explanation for the observed decline. The estimated odds of detected cocaine use reported at the bottom of Table 3 are for the reference population, persons whose attributes matched the reference levels for each attribute analyzed (identified in italics). These odds are referred to as the "reference rate." Hence, urine tests of youthful, male, Black arrestees for a larceny misdemeanor, interviewed by Drug Use Forecasting–Manhattan in 1Q87 and born in 1967, were nine times as likely to give evidence of cocaine use as not. (These odds of 9:1 correspond to a 90% probability of detected cocaine use.)

Table 3 also reports estimated odds ratios associated with each attribute level that differs from the reference level. The odds of detected cocaine use for nonmembers of the reference population equals the product of the reference rate times the odds ratio associated with each attribute differing from the reference level. For example, the analysis indicates that arrestees for drug possession whose other attribute values matched the reference levels were about twice as likely as the reference population to be detected as cocaine users. Their estimated odds are 18:1 (2.04×9).

The Wald statistics reported in Table 3 test whether the variation in estimated odds ratios across the levels of each

TABLE 3—Covariates of Detected Cocaine Use among Youthful Arrestees (n = 1113)[a] (Logistic Regression)

Attribute	Level	Estimated Multiplier
Interview year Wald(6) = 5.5	*1987*[b]	1.00
	1988	0.64
	1989	0.67
	1990	0.54
	1991	0.55
	1992	0.50
	1Q93	0.29
Quarter Wald(3) = 13.9*	*First*	1.00
	Second	0.53
	Third	0.57
	Fourth	0.51
Birth year Wald(11) = 58.1*	*1967*	1.00
	1968	1.73
	1969	0.86
	1970	0.87
	1971	0.55
	1972	0.23
	1973	0.28
	1974	0.15
	1975	0.10
	1976	0.12
	1977	[c]
	1978	[c]
Arrest charge Wald(8) = 27.6*	Drug possession	2.04
	Drug sales	1.76
	Robbery	0.81
	Burglary	1.17
	Larceny/auto theft	1.00
	Violent index	0.60
	Other income generating	1.44
	Other serious crime	0.61
	Other	0.62
Misdemeanor/felony Wald(1) = 16.5*	*Misdemeanor*	1.00
	Felony	0.41
Sex Wald(1) = 1.8	*Male*	1.00
	Female	0.76
Race/ethnicity Wald(3) = 5.6	*Black*	1.00
	White	1.27
	Hispanic	0.97
	Other/missing data	0.41
Odds of detected cocaine use for the reference population = 9:1		

Note. $-2 \times$ log-likelihood = 1213.6.
[a]Excludes three cases that were missing demographic data.
[b]Italics indicate reference level.
[c]Too few cases to accurately estimate a coefficient.
*Statistically significant $\alpha = .01$ level.

TABLE 4—Postdicted Decline in Detected Cocaine Use, 1987 through 1993, Based on Changes in Arrest Sample Composition and Birth Cohorts

Model	Percentage Detected as Cocaine Users (SE) by Interview Year						
	1987	1988	1989	1990	1991	1992	1Q93
Observed	69 (4)	61 (4)	51 (3)	31 (3)	25 (3)	23 (4)	17 (6)
Postdicted by composition and birth cohort differences	66	61	49	33	26	23	19
Postdicted by sample composition differences alone	45	46	43	36	39	40	39
Postdicted by birth cohort differences alone	67	59	49	36	25	21	18

attribute are statistically significant, based on an asymptotic χ^2 distribution. The Wald statistics also provide a rough rank ordering of the partial influence of each independent variable, controlling for all others. The most important variable was birth year, followed by arrest charge, misdemeanor/felony, and quarter. The variation in detected cocaine use associated with interview year was not statistically significant ($\alpha = .01$ level), suggesting that the observed decline probably did not result from a reduction or termination of cocaine use.

Youthful arrestees born in 1968 exhibited the highest rate of detected cocaine use, controlling for the influence of all other variables. These youths reached age 18 in 1986, which was near the height of the epidemic-like growth in the prevalence of crack use.[25–27] The rate of detected cocaine use declined steadily among arrestees born after 1968. Those born in 1976 were 14 times less likely to be detected as cocaine users than those born in 1968 (1.73/0.12), controlling for all other variables.

This finding helps explain the observed decline in detected cocaine use among youthful arrestees, given that the 1968 birth cohort comprised a declining proportion of the sample over time. Since the 1968 birth cohort turned 21 in 1989, the 1989 sample contained fewer arrestees born in 1968—they were included in the youthful arrestee sample when an arrest preceded the offender's 21st birthday—and samples from 1990 and later contained no one born in 1968. Analysis of quarterly data not presented here indicates that the highest levels of detected cocaine use among youthful arrestees occurred in mid-1989 and decreased dramatically in the fourth quarter of that year.

Table 4 presents results from various postdiction models of the decline in detected cocaine use as a function of various independent variables. The first row shows the observed decline in detected cocaine use from 1987 to 1Q93, for comparison. The second row shows time variation in detected cocaine use postdicted by the combined effects of changes in sample composition and differences across birth cohorts. The combined model closely recreates the observed decline, corroborating the finding from the logistic regression analysis, which suggests that the overall decline does not reflect a reduction or termination of cocaine use.

The last two rows of Table 4 show the decline attributable to changes in sample composition alone and birth cohort alone, respectively. Sample composition alone postdicted a modest decline from 45% in 1987 down to 39% in 1Q93, which is much smaller than the observed decline from 69% to 17%. Hence, changes in policing priorities and/or Drug Use Forecasting practices as measured through the individual attributes included in this analysis can account for, at most, 12% [(45%–39%) ÷ (69%–17%)] of the observed decline in detected cocaine use from 1987 to 1Q93.

On the other hand, the decline postdicted by differences across birth cohorts alone recreates the observed decline as well as the model, including sample composition variables and birth cohort. This further supports the idea that the decline in detected cocaine use resulted primarily from differences in the proportion of youthful arrestees initiating use of cocaine across birth cohorts, and was not owing to changes in sample composition or in cocaine use patterns among members of earlier birth cohorts.

Discussion

Three different statistical analyses with the Drug Use Forecasting–Manhattan data—age-period-cohort analysis, lo-

gistic regression, and postdiction—all strongly suggest that the observed decline in cocaine use among youthful arrestees resulted from a spectacular decline in the proportion of persons who grew up more recently becoming habitual users of cocaine or crack. Arrestees born in 1968 reached age 18 in 1986, the height of the crack epidemic,[26,27] and were the most likely to be detected as cocaine users (78%). Detected cocaine use declined sharply among eight subsequent cohorts of youthful arrestees to 45% among those born in 1971 and a low of 10% among those born in 1975 and 1976. Furthermore, cohorts of arrested youths who initiated cocaine use during the height of the cocaine/crack epidemic during the mid-1980s tend to have continued regular use. Whether new cohorts of youthful arrestees, those born between 1976 and 1979, continue to avoid cocaine and crack before reaching age 21 (and at older ages as well) in 1997 to 2000 remains to be documented. Additional research is also needed to document whether parallel declines in detected cocaine use have been occurring among youthful arrestees in other Drug Use Forecasting cities.

By way of explanation, some ethnographic research suggests that youths reaching their teens in the late 1980s and 1990s saw the ravages of crack smoking and heroin injecting by adults and consciously chose not to use these drugs.[21,28] Alternatively, this dramatic decline may represent a major and important victory for the various efforts such as school-based drug-abuse prevention programs and street sweeps of drug dealers aimed at reducing cocaine abuse in inner-city New York. Additional research is needed to identify the role (if any) of various policy initiatives. Identifying the nature of the most effective programs could help inform future drug abuse prevention policy.

The finding that high-risk, inner-city youths reaching age 18 in the 1990s were dramatically less likely to initiate regular use of cocaine suggests that the cocaine and crack epidemics are ebbing. Widespread use of any particular drug cannot be sustained indefinitely unless large numbers of youths continue to become habitual users. The use of cocaine, and in particular crack, may persist for another 20 or 30 years if those who began using it in the mid-1980s continue regular consumption. Perhaps New York City's crack epidemic will parallel the city's experience with heroin injection, initiation to which was quite common during the 1960s and early 1970s but is undertaken by relatively few young people today.[9] Many persons who began injecting heroin during its period of widest popularity (1963 to 1973) are still injecting it even though they tend to be in their 30s, 40s, and 50s in 1994. An analogous trajectory for crack cocaine would result in substantial numbers of middle-aged crackheads but few youthful crack users in the 2010s. □

Acknowledgments

This research was primarily supported by a grant from the National Institute of Justice on "Expanding Applications of DUF Data in New York City" (1-7502-NY-IJ) and by a grant from the National Institute on Drug Abuse on "The Natural History of Crack Distribution" (1 R01 DA05126-04). Additional support was provided by National Development and Research Institutes, Inc (formerly Narcotic and Drug Research, Inc).

References

1. Kandel DB. Convergence in prospective longitudinal surveys of drug use in normal populations. In: Kandel DB, ed. *Longitudinal Research on Drug Use: Empirical Findings and Methodological Issues*. Washington, DC: Hemisphere Publishing; 1978:3–38.
2. Johnston LD. Toward a theory of drug epidemics. In: Donohew DH, Sypher H, Bukoski WJ, eds. *Persuasive Communication and Drug Abuse Prevention*. Hillsdale, NJ: Lawrence Erlbaum; 1991:93–132.
3. Kandel D, Yamaguchi K. From beer to crack: developmental patterns of drug involvement. *Am J Public Health*. 1993;83: 851–855.
4. Golub A, Johnson BD. The shifting importance of alcohol and marijuana as gateway substances among serious drug abusers. *J Stud Alcohol*. 1994;55:607–614.
5. Hunt LG, Chambers CD. *The Heroin Epidemics: A Study of Heroin Use in the US, 1965–1975 (pt 2)*. Holliswood, NY: Spectrum; 1976.
6. Boyle J, Brunswick AF. What happened in Harlem? Analysis of a decline in heroin use among a generation unit of urban black youth. *J Drug Issues*. 1980;10:109–130.
7. Clayton RR, Voss HL. *Young Men and Drugs in Manhattan: A Causal Analysis*. Rockville, Md: National Institute on Drug Abuse; 1981. National Institute of Drug Abuse Monograph.
8. Johnson BD, Manwar A. Towards a paradigm of drug eras. Presentation at the American Society of Criminology annual meeting; November 1991; San Francisco, Calif.
9. Golub A, Johnson BD. Cohort differences in drug use pathways to crack among current crack abusers in New York City. Presentation at the American Society of Criminology annual meeting; November 1992; New Orleans, La.
10. Johnson BD, Golub A, Hossain M. Trends in heroin use among arrestees in the Drug Use Forecasting program. Presented at the American Society of Criminology annual meeting; November 1992; New Orleans, La.
11. Johnston LD, O'Malley PM, Bachman JG. Press releases from the University of Michigan (concerning Monitoring the Future). April 9, 1993, and January 27, 1994.
12. *National Household Survey on Drug Abuse: Main Findings 1990*. Rockville, Md: National Institute on Drug Abuse; 1991. DHHS publication ADM 91-1788.
13. *Preliminary Estimates from the 1992 National Household Survey on Drug Abuse*. Rockville, Md: National Institute on Drug Abuse; 1993.
14. National Institute of Justice. Drug Use Forecasting (DUF) research update. *Research in Brief*, NCJ 134230. July 1992.
15. Wish ED, Gropper BA. Drug testing by the criminal justice system: methods, research and application. In: Tonry M, Wilson JQ, eds. *Drugs and Crime*. Chicago, Ill: University of Chicago Press; 1990;13:321–391. Crime and Justice Series.
16. Chaiken JM, Chaiken MR, Poulin ER. Understanding the Drug Use Forecasting (DUF) sample of adult arrestees. Report to the National Institute of Justice. Lincoln, Mass: LINC; 1993.
17. Drug use measurement: strengths, limitations, and recommendations for improvement. Report to the Chairman, Committee on Government Operations. Washington, DC: US General Accounting Office; June 1993. GAO/PEMD-93-18.
18. Kandel DB, Davies M. Decline in the use of illicit drugs by high school students in New York State: a comparison with national data. *Am J Public Health*. 1991;81: 1064–1067.
19. Williams T. *The Cocaine Kids: The Inside Story of a Teenage Drug Ring*. Reading, Mass: Addison-Wesley; 1989.
20. Johnson BD, Williams T, Dei K, Sanabria H. Drug abuse and the inner city: impact on hard drug users and the community. In: Tonry M, Wilson JQ, eds. *Drugs and Crime*. Chicago, Ill: University of Chicago Press; 1990;13:9–67. Crime and Justice Series.
21. Hamid A. The developmental cycle of a drug epidemic: the cocaine smoking epidemic of 1981–1991. *J Psychoactive Drugs*. 1992;24:337–348.
22. Williams T. *Crackhouse*. Reading, Mass: Addison-Wesley; 1992.
23. Golub A, Johnson BD. Modeling the epidemic of onset to crack abuse. Presentation at the annual meeting of the Society for the Study of Social Problems; August 1992; Pittsburgh, Pa.
24. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NY: John Wiley and Sons; 1989.
25. Johnson BD. The crack era in New York City. *Addict Recovery*. 1991;May/June:24–27.
26. Johnson BD, Lewis C, Golub A. Crack onset in the 1980s in New York City. In: Vamos P, Corriveau PJ, eds. *Drugs and Society to the Year 2000: Proceedings of the XIV World Congress of Therapeutic Communities*. Montreal, Canada: Portage Program for Drug Dependencies Inc; 1992:365–369.
27. Golub A, Johnson BD. Modeling the onset to crack abuse. Presentation at the annual meeting of the Society for the Study of Social Problems; August 1992; Pittsburgh, Pa.
28. Dunlap E, Johnson BD. The setting for the crack era: macro forces, micro consequences 1960–92. *J Psychoactive Drugs*. 1992;24:307–322.

Cognitive Impairment and Mortality in Older Community Residents

ABSTRACT

Objectives. Cognitive impairment among the elderly has been linked to mortality in studies of clinical populations. The purpose of this study was to examine the mortality risk associated with cognitive impairment among elderly populations in the community.

Methods. Cognitive impairment and other social and health factors were assessed in 1855 elderly community residents. This sample was reinterviewed periodically to assess changes in health and survival.

Results. At baseline 33% of the sample were mildly impaired and 8% were severely impaired. Across a 48-month observation period the survival probability was .85 for the cognitively unimpaired, .69 for the mildly impaired, and .51 for severely impaired respondents. When adjustments were made for the effects of other health and social covariates, severely impaired persons were twice as likely to die as unimpaired persons. Those who were mildly impaired were also at an increased risk.

Conclusions. Other investigators have found that cognitive impairment is a significant predictor of dementia. We found that it is a significant predictor of mortality as well. Early detection of impaired cognition and attention to associated health problems could improve the quality of life of these older adults and perhaps extend their survival. (*Am J Public Health.* 1994;84:1255–1260)

Howard R. Kelman, PhD, Cynthia Thomas, PhD, Gary J. Kennedy, MD, and Jiming Cheng, PhD

Introduction

Deficits in cognitive functioning, prevalent among the elderly, are indicators of disease and disability.[1,2] Recent studies among noninstitutionalized populations, employing brief screening tests, indicate that 20% or more of the elderly population may be cognitively impaired. Three or four times as many persons have mild or moderate impairment.[3–5] There is a greater prevalence of cognitive impairment among the older old, those with less education, and minorities.[6–8]

Cognitive impairment is a threat to the quality of life of individuals and their families,[9,10] contributes to increased hospital use,[11,12] and is a risk factor for institutionalization.[13,14] The prevalence of shortened survival among those with dementing disorders[15,16] raises a question about whether milder forms of impairment are associated with mortality. The availability of brief yet reliable measures of cognitive performance[17,18] facilitates such studies.

Most studies linking mortality to cognitive impairment, however,[1] are based on small numbers of elderly individuals with dementing disorders. These studies have employed various measures of cognitive impairment, including detailed clinical assessments. Survival rates have varied by severity of impairment, age, sex, and comorbidity.[19–24]

The few studies of elderly nondemented community residents differ in the representativeness of the samples, measures of cognitive impairment, length of follow-up, and whether other predictors of mortality were controlled.

Two longitudinal studies of nonrepresentative samples of elderly community residents, one of aged twins,[25] the other of very old persons,[26] yielded contrasting findings with regard to the linkage of mortality, impairment, and advancing age. No linkage was found in the study of twins. Increased mortality was associated with age in the study of very old persons later diagnosed as demented.

In a Scottish general practice population, a relative risk of early death of 3.5 was found for elderly cognitively impaired persons compared with nonimpaired persons, but there was no difference in risk of death between those with severe and moderate impairment.[26] These findings resemble those of Jorm et al. in an Australian community sample.[27] Neither study, however, employed multivariate analysis to control for potential risk factors other than age.

Shapiro and Tate recently estimated relative risks from Cox proportional hazards models for a subsample of participants in the Manitoba Longitudinal Study of Aging.[28] In addition to age and sex, cognitive impairment and a diagnosis of dementia contributed significantly to mortality. Potential risk factors including illness, income, and education were not studied, however, possibly because data

Howard R. Kelman, Cynthia Thomas, and Jiming Chen are with the Department of Epidemiology and Social Medicine, Montefiore Medical Center and Albert Einstein College of Medicine, Bronx, NY. Cynthia Thomas is also with Westat Inc, Rockville, Md, and Jiming Chen is also with the Columbia University Center for Geriatrics and Gerontology, New York, NY. Gary J. Kennedy is with the Department of Psychiatry, Montefiore Medical Center and Albert Einstein College of Medicine.

Requests for reprints should be sent to Howard R. Kelman, PhD, Department of Epidemiology and Social Medicine, Montefiore Medical Center, 111 E 210th St, Bronx, NY 10467.

This paper was accepted December 28, 1993.

TABLE 1—Selected Characteristics of the Study Sample (n = 1855)

Female, %	66.3
Age, %	
65–74 y	48.8
75–84 y	40.6
85+ y	10.7
Married, %	40.7
One-person household, %	43.8
Race other than White, %	5.0
Hispanic ethnicity, %	5.6
Employed full- or part-time, %	10.3
Mean annual income, $	12 039
Medicaid, %	14.6
Private health insurance, %	68.1

Note. Percentages, means, standard deviations, and all other calculations in this and subsequent tables are based on weighted data; totals are unweighted. Weights were applied to adjust for unequal probability sampling.

were limited to physician and hospital records.

Liu et al., studying a subsample from the Framingham Health Study,[29] related performance on a battery of neuropsychological tests to mortality over an 8- to 10-year period. Poor test performance was associated with an increased risk of death even after adjustment for age, education, and illness. The generalizability of these findings may be limited because of the study's high and selective nonparticipation rates. The validity of the test battery as a measure of cognitive impairment is also uncertain.

In this paper we report findings on the association between cognitive impairment and all-cause mortality in a large sample of urban elderly Medicare recipients.[30] Baseline measures of cognitive impairment were assessed as mortality risk factors by means of both univariate and multivariate statistical procedures, with adjustment for the effects of potentially confounding demographic, health, and social variables.

Methods

Sample

Data for this analysis are from personal interviews with 1855 community residents at least 65 years of age who were participants in a longitudinal study of aging and health (Norwood-Montefiore Aging Study [NMAS]).[31] The sample was randomly selected from a list of Medicare beneficiaries living in a neighborhood in the Bronx. Characteristics of the study sample are shown in Table 1. Two thirds were women, more than half were 75 years of age or older, 41% were married, and 44% lived alone. The mean annual income in 1984 was $12 039. Two thirds had private health insurance in addition to Medicare, and nearly 15% received Medicaid. The small proportions of non-Whites and Hispanics reflect the pattern of housing segregation in the study area.

Interviews conducted between July 1984 and March 1985 were the first in a series of semiannual contacts over 3½ years. The response rate to the baseline interview was 73%. Ninety-five percent or more of baseline respondents were reinterviewed in each successive interview wave. Dates and circumstances of death were obtained from proxies, usually family members. Respondents who discontinued participation were recontacted to ascertain survival status. Data were weighted to equalize the representation of elderly people living in households with more than one eligible person.

A 50% sample of nonrespondents was resurveyed 18 months after the health study was under way. Eighty-six percent of these persons, or their proxies, were recontacted. Compared with respondents, more nonrespondents assessed their health as poor or fair. They were more often women, were poorer, and were more likely to live alone. They were not older, however, and interviewers rated nearly equal proportions of both groups as "confused" (7.6% vs 6.6%).[32]

Measures

The Mini-Mental State Exam (MMSE), an instrument widely used to screen for cognitive impairment, was administered at baseline. Developed by Folstein et al., the MMSE assesses signs of dysfunction in orientation, registration, attention, calculation, recall, and language.[33] The MMSE was originally validated with hospitalized patients and subsequently adapted for use in the National Institute of Mental Health's Epidemiological Catchment Area Program.[34] It has been widely employed in studies in this country and abroad.[35–37]

MMSE scores range from 0 to 30. Persons with scores of less than 18 are defined as severely impaired, those with scores of 18 to 23 as mildly impaired, and those with scores of 24 and above as unimpaired. Although the MMSE is an easily administered and reliable tool to identify signs of cognitive impairment, scores do not provide clinical diagnoses.[38] Cognitive impairment may result from a variety of mental conditions including dementia, depression, brain injury, and mental retardation.[39]

Potentially significant mortality risk factors identified in the literature[40–42] were also measured at baseline. Demographic characteristics included age (measured in 5-year groups), sex, and race. Social characteristics included marital status, income (measured in $1000 increments), education (0–8 years vs more), living arrangement (alone or with others), whether any language in addition to English was spoken at home, and receipt of social support (both formal and informal vs neither or only one). Health characteristics included self-assessed health (fair/poor vs good/excellent), number of problems in activities of daily living, number of medical conditions, presence of two or more cardiovascular conditions, symptoms of depression (a Center for Epidemiologic Studies–Depression [CES–D] score less than 18), and visual or hearing problems.

Most of the items defining these variables were constructed from questions taken from the Aging Supplement of the National Health Interview Survey (NHIS) and the National Medical Care Expenditure Survey. Measures of social support and problems in daily living were modified to suit the character of the study sample. Social support was measured according to the frequency and amount of assistance received from family members (informal) or provided by agencies (formal) for personal care, shopping, traveling, money management, and the like, closely paralleling similar items employed by other researchers.[43] Problems in daily living (difficulties in carrying out daily activities) were similar to those used in other current research on functional performance.[44]

Analysis

The analytic strategy was survival analysis, employing Cox proportional hazards models in which the dependent variable, time to death, is predicted by the independent variables and by the MMSE score. This procedure uses all available information about study participants as long as they are followed and is therefore applicable when there are censored observations.

The validity of the proportional hazards assumption was assessed for each predictor by testing whether or not the hazard ratio changed with time. We found

no evidence of an increasing or decreasing trend in the hazard ratios.

Survival curves, by level of cognitive impairment at baseline, were constructed by means of a Kaplan-Meier product limit estimator and tested for homogeneity with the log rank test. The relationship between the independent variables and survival time was assessed with univariate Cox proportional hazards models. Statistically significant variables ($P < .05$) in the univariate analyses were incorporated as predictors in multivariate Cox regression models. Analyses were performed with the fast backward elimination procedure and with predictor variables entered simultaneously. Mortality risk factors for each cognitive impairment subgroup were identified by univariate and multivariate Cox regression analyses. Demographic, health, and social characteristics of the unimpaired and the two impaired subgroups were also compared. Differences among the groups were tested for significance with *t* tests (for continuous variables) or the chi-square test (for dichotomous variables).

Results

Characteristics of Cognitive Impairment Groups

At baseline, 41% of respondents scored as impaired: 33% scored in the mild range (MMSE scores 18–23) and 8% in the severely impaired category (MMSE scores <18). These rates, particularly for mild impairment, are higher than those found in the New Haven and Baltimore Epidemiological Catchment Area sites.[3,4] However, differences in scoring the MMSE (the Epidemiological Catchment Area scoring procedures modified Folstein's original scoring instructions), in the age distributions of the samples, and in administration of the MMSE may account in part for these different rates. The prevalence of impairment for men (39.2%) was not significantly different from that of women (42.1%). Advancing age was associated with severity of impairment.

Health and social characteristics of the unimpaired, mildly impaired, and severely impaired groups at baseline are shown in Table 2. Severely impaired persons compared with their unimpaired peers were older, had more health problems, were poorer, were not as well educated, and were more likely to be women.

Mildly impaired persons differed significantly from their unimpaired peers in age, income, and other socioeconomic characteristics, but not in level of education or in most measures of disability and health. Mildly impaired respondents were younger than their severely impaired peers, tended to be men, had higher levels of income and education, were more often married, less often lived alone, and reported fewer health problems.

TABLE 2—Baseline Characteristics of the Study Sample, by Level of Cognitive Impairment

	Level of Impairment			
	A Severe (n = 168)[a]	B Mild (n = 746)[a]	C No Impairment (n = 1340)[a]	Total Sample (n = 2254)[c]
Sociodemographic characteristics				
Mean age, y (SD)	80.5[C] (7.5)	76.4[C] (6.9)	73.8 (5.8)	75.4 (6.6)
Female, %	73[B]	66	65 (10 548)	66.3
Mean annual income $ (SD)	7179 (5477)	9763[A] (7648)	13 902[A,B] (10 458)	12 039 (9 512)
Education >8 y, %	28	44[A]	71[A,B]	58.9
Married, %	25.7[B,C]	41.5[A]	42.2[A]	40.7
Non-White, %	8	7	3[A,B]	5.1
Receiving social support, %	27.4[B,C]	12.2[A,C]	7.7[A,B]	11.4
Living alone, %	54.8[B,C]	44.6[A]	42.0[A]	43.8
Language,[b] %	67.6[B,C]	50.1[A,C]	43.0[A,B]	47.6
Health characteristics				
Fair/poor self-assessed health, %	53[B,C]	42	40	41.3
Mean no. of problems in daily living[c]	2.0[B,C] (2.5)	0.9 (1.7)	0.5 (1.2)	0.7 (1.6)
Depressive symptoms, %	26[B,C]	15	13	14.4
Vision problems, %	41[C]	35	32	34.7
Hearing problems, %	46[C]	42[C]	34	37.8
Mean no. of medical conditions (SD)	2.4[B] (1.9)	2.1 (1.5)	2.3[B] (1.7)	2.2 (1.6)
Arthritis, %	58[B,C]	50	50	51.0
Stroke, %	8[C]	7[C]	4	5.1

Note. Table entries accompanied by capital letters in superscript are significantly greater than values for groups represented by the column letter at $P < .05$ by chi-square or *t* test.
[a]Weighted totals.
[b]Language other than English spoken at home in addition to English.
[c]Problems in performing daily activities.

Survival Probabilities

Sixteen percent of the sample died during a 4-year observation period. These persons represented 11% of the unimpaired, 20% of the mildly impaired, and 38% of the severely impaired persons.

Survival curves for impairment subgroups are shown in Figure 1. Probabilities of survival for severely impaired persons (.51), mildly impaired persons (.69), and unimpaired persons (.85) were all significantly different.

Predictors of Mortality (Time to Death)

The relationship between time to death and each baseline predictor variable, including severe and mild cognitive impairment, was assessed initially by means of a Cox univariate proportional hazards model. Results are shown in Table 3. Regression coefficients for nearly all demographic, social, and health variables were statistically significant. Except for income, each manifested a significant relative risk for shortened survival.

Statistically significant variables were next entered into a multivariate Cox regression model to assess the indepen-

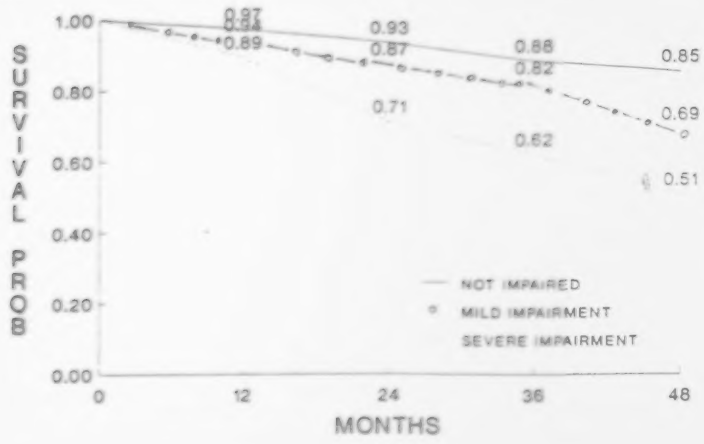


FIGURE 1—Survival curves and probabilities of survival, by level of cognitive impairment.

TABLE 3—Univariate Predictors of Mortality (Time to Death)

	Rate Ratio	95% CI
Age[a]	1.5	1.38, 1.62
Sex[b]	1.7	1.39, 2.13
Income	1.0	1.02, 1.06
Education[c]	1.6	1.27, 1.92
Problems in daily living	1.3	1.21, 1.32
Fair/poor self-assessed health	1.9	1.52, 2.29
Two or more cardiovascular conditions	1.7	1.27, 2.16
Depression	1.5	1.15, 1.91
Visual problem	1.3	1.02, 1.55
Hearing problem	1.5	1.26, 1.89
Receiving social support	2.9	2.26, 3.67
Married[d]	1.3	1.12, 1.56
Living alone	NS	. . .
Language	NS	. . .
Cognitive impairment		
Severe	4.2	3.10, 5.59
Mild	1.9	1.48, 2.33

Note. NS = not significant; CI = confidence interval.
[a]Categorized in sequential 5-year age groups (65–69, 70–74, etc., to 85+ years).
[b]Rate ratio calculated for males relative to females.
[c]Rate ratio calculated for those with less than 8 years of education relative to those with more education.
[d]Rate ratio calculated for those other than married relative to those married.

TABLE 4—Multivariate Predictors of Mortality: Total Sample (n = 1855)

	Rate Ratio	95% CI
Advancing age	1.3	1.16, 1.38
Low income	1.1	1.01, 1.05
Problems in daily activities	1.1	1.03, 1.19
Fair/poor self-assessed health	1.6	1.12, 1.84
Receiving social support	1.5	1.16, 1.99
Sex: male	1.1	1.18, 1.52
Severe impairment	2.2	1.13, 2.69
Mild impairment	1.5	1.12, 1.82
Education	NS	. . .
Married	NS	. . .
Depression	NS	. . .
Two or more cardiovascular conditions	NS	. . .

Note. NS = not significant; CI = confidence interval.

dent contribution of each variable, controlling for the effects of others. Results of the backward elimination procedure are shown in Table 4. The risk ratio was 2.2 for severe impairment and 1.5 for mild impairment. Thus, even when health, demographic, and social factors are controlled, cognitive impairment whether severe or mild, remains an important mortality risk. Other predictors were male sex, receipt of both formal and informal social support, fair or poor self-assessed health, two or more cardiovascular conditions, and advancing age.

Because men and women in this sample differed by marital status and other health and social characteristics, we entered interaction terms for sex and key independent variables in the multivariate models. However, high correlations between these variables and the interaction terms—for example, .59 (age), .76 (problems in daily living), .75 (self-assessed health), and .85 (social support)—precluded their inclusion in the model. We therefore developed multivariate models separately for men and for women. The results of the models, in which a backward elimination procedure was used, are shown in Table 5.

For both men and women, cognitive impairment is a significant risk factor; the risk is 2.0 for men and 2.3 for women for severe impairment. For mild impairment, the risk for both men and women was 1.5. Additional risk factors for men were poor self-assessed health, more problems in daily living, and low income. Additional risk factors for women were receipt of social support and more problems in daily living. Advancing age was a risk factor common to both.

Risk factors predicting mortality for the entire sample or for each sex, however, may differ by level of cognitive impairment, particularly given the differences in health and social characteristics and in mortality risk associated with different levels of impairment. To explore this issue, additional univariate and multivariate regression analyses were performed for each cognitive status subgroup (unimpaired, mildly impaired, and severely impaired). Results of the multivariate analyses, which employed backward elimination procedures, are shown in Table 6.

Advancing age is the only predictor common to all three groups. In addition to age, receipt of both formal and informal social support (a possible proxy for dependency) and only English spoken at home predicted shortened survival for severely impaired respondents. Absent in this subgroup are other health and social predicators.

Among the mildly impaired, less education, poorer health and functional status (more problems in daily living), receipt of both formal and informal support, advancing age, and male sex were risk factors. These same risk factors, except for social support, low education,

TABLE 5—Multivariate Predictors of Mortality, by Sex

	Rate Ratio	95% CI
Men (n = 631)		
Age	1.3	1.13, 1.42
Income	1.0	1.01, 1.05
Problems in daily activities	1.2	1.07, 1.27
Fair/poor self-assessed health	1.8	1.27, 2.43
Social support	NS	. . .
Education	NS	. . .
Depression	NS	. . .
Hearing	NS	. . .
Two or more cardiovascular conditions	NS	. . .
Severe impairment	2.0	1.13, 3.23
Mild impairment	1.5	1.07, 2.10
Women (n = 1223)		
Age	1.3	1.13, 1.45
Income	NS	. . .
Problems in daily activities	1.1	1.05, 1.21
Fair/poor self-assessed health	NS	. . .
Social support	1.8	1.22, 2.52
Education	NS	. . .
Married	NS	. . .
Race (Non-White)	NS	. . .
Depression	NS	. . .
Hearing	NS	. . .
Two or more cardiovascular conditions	NS	. . .
Severe impairment	2.3	1.52, 3.59
Mild impairment	1.5	1.07, 2.06

Note. NS = not significant; CI = confidence interval.

TABLE 6—Multivariate Predictors of Mortality, by Cognitive Impairment Status

	No Impairment		Mild/Moderate Impairment		Severe Impairment	
	RR	95% CI	RR	95% CI	RR	95% CI
Advancing age	1.4	1.20, 1.58	1.1	1.00, 1.29	1.3	1.04, 1.63
Sex: male	2.7	1.89, 3.72	2.6	1.81, 3.57	NS	. . .
Education	NS	. . .	1.6	1.15, 2.27	NS	. . .
Language	NS	. . .	NS	. . .	2.0	1.18, 3.27
Married	1.6	1.13, 2.33	NS	. . .	NS	. . .
Fair/poor self-assessed health	1.8	1.24, 2.47	1.6	1.12, 2.23	NS	. . .
Problems in daily activities	1.2	1.13, 1.34	1.1	1.04, 1.24	NS	. . .
Receiving social support	NS	. . .	1.9	1.24, 3.00	1.8	1.06, 3.12
Two or more cardiovascular conditions	NS	. . .	1.5	1.06, 2.18	NS	. . .

Note. RR = rate ratio; CI = confidence interval; NS = not significant.

and presence of two or more cardiovascular conditions, apply to the unimpaired subgroup.

Discussion

These results indicate that signs of cognitive impairment, both mild and severe, represent a significant threat to the survival of elderly community residents. In this sample, severely impaired persons were more than twice as likely as unimpaired persons to die within 4 years. Mildly impaired persons also had an increased risk for shortened survival, consistent with a recent finding from a Swedish sample.[35] Mortality risks were associated with severe and mild cognitive impairment even when adjustments were made for health and social factors, and the association was evident for men as well as women.

These findings, based on a large community sample, support the link between cognitive impairment and mortality found in the few studies of elderly persons without a diagnosis of dementia. These findings also point to the importance for shortened survival of mild cognitive impairment, a widely prevalent level of dysfunction among elderly persons in the community. Perhaps because there is often no obvious link to a known medical condition, mild impairment has not often been examined in studies of mortality, nor have physicians and others who provide health services to the elderly been encouraged to screen for its presence.[45]

Cognitive impairment may contribute to mortality in different ways for severely and mildly impaired individuals. Its progression may be a marker for the decline in health of those with dementing illness. Mild cognitive impairment may have a more indirect influence on mortality.[46] Deficiencies in language comprehension, in recall, or in other cognitive areas may contribute to failure to seek timely health care, to use recommended treatments or medications, or to recognize signs and symptoms of disease. If cognitive deficits limit educational and vocational achievement, the resulting socioeconomic disadvantage may then lead to poorer health and increased mortality.

Although this study was based on a large and randomly selected sample, the findings may be limited by the characteristics of the population. Compared with 1984 estimates of the US elderly population, the study sample was older, less affluent, and contained more women and more persons living alone. (Comparison data are available on request.) Data on mortality were based on proxy reports and not death certificates. However, information was obtained within 6 months of death, limiting the potential for recall error. Finally, cognitive impairment was measured by a screening instrument rather than a clinical assessment. Despite these potential limitations, the results are consistent with findings from the Manitoba and Framingham studies, even though the studies differ in samples and in measures of cognitive impairment.

Severe and mild impairment may be different conditions with different origins or different points along a continuum of cognitive decline. Although our data did not permit us to investigate this issue, some individuals with mild signs of cognitive dysfunction may have had a disadvantaged social status or poor health early on that may have limited the expression of cognitive capacity or even stunted its development. On the other hand, mild cognitive impairment may also herald the emergence of dementing illness among those who are not disadvantaged. Strategies designed to lessen the impact of cognitive impairment on the functioning of elderly people require diagnosis of the reasons for the impairment.

The prevalence of cognitive impairment among the elderly, particularly mild impairment, coupled with forecasts for increasing numbers of older persons into the next century, points to the need to train health care professionals to recognize this disability and deal with its potentially negative health consequences. □

Acknowledgments

The work upon which this study is based was supported by grants P01 AGO 3424 and R01 AGO 8125 from the National Institute on Aging (Howard R. Kelman, PhD, principal investigator).

A preliminary version of this paper was presented at the annual meeting of the Gerontological Society of America, San Francisco, Calif, November 23, 1991.

We are indebted to David Biderman, BA, for research assistance and to Ms Edie Spinelli for able secretarial assistance.

References

1. Kasper JD. Cognitive impairment among functionally limited elderly people in the community: future considerations for long-term care policy. *Milbank Q.* 1990;68:81–109.
2. Spector WD. Cognitive impairment and disruptive behavior among community-based elderly persons: implications for targeting long-term care. *Gerontologist.* 1991;31:51–59.
3. Holzer CE III, Tischler GL, Leaf PJ, Myers JK. An epidemiological assessment of cognitive impairment in a community population. In: Greenley JR, ed. *Research in Community and Mental Health.* Greenwich, Conn: JAI Press; 1984;4:3–32.
4. Folstein M, Anthony JC, Parhad I, Duffy B, Greenberg EM. The meaning of cognitive impairment in the elderly. *J Am Geriatr Soc.* 1985;33:228–235.
5. Escobar JI, Burman A, Karno M, Forsythe A, Landsverk J, Golding JM. Use of the Mini-Mental State Examination (MMSE) in a community population of mixed ethnicity. *J Nerv Ment Dis.* 1986;174:607–614.
6. Jorm AF, Scott R, Henderson AS, Kay DWK. Educational level differences on the Mini-Mental State Examination: the role of test bias. *Psychol Med.* 1988;18:727–731.
7. Fillenbaum G, Hughes DC, Heyman A, George LK, Blazer DG. Relationship of health and demographic characteristics to Mini-Mental State Examination score among community residents. *Psychol Med.* 1988;18:719–726.
8. Murden RA, McRae TD, Kaner S, Bucknam ME. Mini-Mental State Examination scores vary with education in blacks and whites. *J Am Geriatr Soc.* 1991;39:149–155.
9. Moritz DJ, Kasl SV, Berkman LF. The health impact of living with a cognitively impaired elderly spouse: depressive symptoms and social functioning. *J Gerontol.* 1989;44:S17–S21.
10. Deimling GT, Bass DM. Symptoms of mental impairment among elderly adults and their effects on family caregivers. *J Gerontol.* 1986;41:778–784.
11. Weiler PG, Lubben JE, Chi I. Cognitive impairment and hospital use. *Am J Public Health.* 1991;81:1153–1157.
12. Binder EF, Robin LN. Cognitive impairment and length of hospital stay in older persons. *J Am Geriatr Soc.* 1990;58:759–766.
13. Branch LG, Jette AM. A prospective study of long-term care institutionalization among the aged. *Am J Public Health.* 1982;72:1373–1379.
14. Kelman HR, Thomas C. Transitions between community and nursing home residence in an urban elderly population. *J Community Health.* 1990;15:105–122.
15. Jorm AF, Korten AE, Henderson AS. The prevalence of dementia: a quantitative integration of the literature. *Acta Psychiatr Scand.* 1987;76:465–479.
16. Evans DA, Funkenstein HH, Albert MS, et al. Prevalence of Alzheimer's disease in a community population of older persons. *JAMA.* 1989;262:2551–2556.
17. Kahn RL, Goldfarb AI, Pollack M. Brief objective measures for the determination of mental status in the aged. *Am J Psychiatry.* 1960;117:326–328.
18. Blum JE, Clark ET, Jarvick LF. The New York State Psychiatric Institute Study of Aging Twins. In: Jarvick JF, Eisdorfer C, Blum JE, eds. *Intellectual Functioning in Adults.* New York, NY: Springer; 1973:13–19.
19. Barclay LL, Zemcov A, Blass JP, Sansone V. Survival in Alzheimer's disease and vascular dementias. *Neurology.* 1985;35:834–840.
20. Heyman A, Wilkinson WE, Hurwitz BJ, et al. Early-onset Alzheimer's disease: clinical predictors of institutionalization and death. *Neurology.* 1987;37:980–984.
21. Walsh JS, Welch G, Larson EB. Survival of outpatients with Alzheimer type dementia. *Ann Intern Med.* 1991;21:363–370.
22. Burns A, Lewis G, Jacoby R, Levy R. Factors affecting survival in Alzheimer's disease. *Psychol Med.* 1991;21:363–370.
23. Kaszniak AW, Fox J, Gardell DL, Garron DC, Huckman MS, Ramsay RG. Predictors of mortality in presenile and senile dementia. *Ann Neurol.* 1978;3:246–252.
24. Van Dijk PTM, Dippel DWJ, Habbema JDF. Survival of patients with dementia. *Ann Geriatr Soc.* 1991;39:603–610.
25. Aronson MK, Ooi WL, Geva DL, Masur D, Blau A, Frishman W. Age-dependent incidence, prevalence and mortality in the old. *Arch Intern Med.* 1991;151:989–992.
26. Eagles JM, Beattie JAG, Rentall DB, Rawlinson F, Hagen S, Ashcroft GW. Relation between cognitive impairment and early death in the elderly. *BMJ.* 1990;300:239–240.
27. Jorm AF, Henderson AS, Kay DWK, Jacomb PA. Mortality in relation to dementia, depression and social integration in an elderly community sample. *Int J Geriatr Psychiatry.* 1991;6:5–11.
28. Shapiro E, Tate RB. The impact of a mental status score and a dementia diagnosis on mortality and institutionalization. *J Aging Health.* 1991;3:28–46.
29. Liu Y, Lacroix AZ, White LR, Kittner SJ, Wolf P. Cognitive impairment and mortality: a study of possible confounders. *Am J Epidemiol.* 1990;132:136–143.
30. Kelman HR. Commentary: Aging, health and health care. *J Community Health.* 1990;15:71–73.
31. Kelman HR, Thomas C. Hospital and ambulatory service use by the urban elderly under different health care delivery systems. *Med Care.* 1988;26:739–749.
32. Thomas C. The effect of nonresponse and attrition on samples of elderly people. Washington, DC: US Dept of Health and Human Services; 1989. DHHS publication PHS 3447.
33. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12:189–198.
34. Robins LN, Helzer E, Weissman MM, et al. Lifetime prevalence of specific psychiatric disorders in three sites. *Arch Gen Psychiatry.* 1984;41:149–958.
35. Johansson B, Zarit SH, Berg S. Changes in cognitive functioning in the oldest old. *J Gerontol.* 1992;47:P75–P80.
36. Park JH, Kwon YC. Modification of the Mini-Mental State Examination for use with the elderly in a non-western society, part 1: Development of Korean version of Mini Mental State Examination. *Int J Geriatr Psychiatry.* 1990;5:381–387.
37. Yu ESH, Liu WT, Levy P, et al. Cognitive impairment among elderly adults in Shanghai, China. *Gerontology.* 1989;3:S97–S107.
38. Tombaugh TN, McIntyre NJ. The Mini-Mental State Examination: a comprehensive review. *J Am Geriatr Soc.* 1992;60:922–935.
39. Anthony C, LeResche L, Niaz U, Von Korff MR, Folstein MF. Limits of the "Mini-Mental State" as a screening test for dementia and delirium among hospital patients. *Psychol Med.* 1982;12:397–408.
40. Manton KC. Changing concepts of morbidity and mortality in the elderly population. *Milbank Memorial Fund Q.* 1982;61:362–396.
41. Rice D, Feldman J. Living longer in the United States: demographic changes and health needs of the elderly. *Milbank Memorial Fund Q.* 1983;61:362–396.
42. Verbruqqe LM. Longer life but worsening health? Trends in health and mortality of middle aged and older persons. *Milbank Memorial Fund Q.* 1984;62:475–519.
43. Kelman HR, Thomas C, Tanaka JS. Longitudinal patterns of social support. *Soc Sci Med.* 1994;38:905–914.
44. Spector WD, Katz S, Murphy JB, Fulton JP. The hierachical relationship between activities of daily living and instrumental activities of daily living. *J Chronic Dis.* 1987;40:481–489.
45. Somerfield MR, Weisman CS, Ury W, Chase GA, Folstein MF. Physician practices in the diagnosis of dementing disorder. *J Am Geriatr Soc.* 1991;39:172–175.
46. Bassett SS, Folstein MF. Cognitive impairment and functional disability in the absence of psychiatric diagnosis. *Psychol Med.* 1991;21:77–84.

The US Economic and Social Costs of Alzheimer's Disease Revisited

ABSTRACT

Objectives. An earlier paper estimated the per-case and national incidence costs of Alzheimer's disease for 1983. This paper updates the estimates of costs per case to 1991 and presents new national prevalence estimates of the economic and social costs of the disease.

Methods. All data for the cost estimates were taken from published sources or provided by other researchers.

Results. At midrange values of the estimated cost and epidemiological parameters, the discounted (at 4%) direct and total costs of Alzheimer's disease were $47 581 and $173 932 per case, respectively. The estimated 1991 national direct and total prevalence costs were $20.6 billion and $67.3 billion, respectively. Assuming conservatively that the prevalence of the disease remains constant, the estimated discounted present values of the direct and total costs of all current and future generations of Alzheimer's patients are $536 billion and $1.75 trillion, respectively.

Conclusions. The $536 billion and $1.75 trillion figures are minimum estimates of the long-term dollar losses to the US economy in 1991 caused by Alzheimer's disease. (*Am J Public Health.* 1994;84:1261–1264)

Richard L. Ernst, PhD, and Joel W. Hay, PhD

Introduction

This paper presents per-case and prevalence estimates of the costs of Alzheimer's disease for the year 1991. The data for the estimates were taken from published sources or provided to us by other researchers. In a previous paper, we estimated the costs of this disease for the year 1983.[1] In this revision we have drawn on a newer and more extensive body of findings.

Methods

The methodology used here is discussed in our earlier paper.[1] Briefly, it is assumed that (1) the cost of Alzheimer's disease is an opportunity cost measured apart from the normal costs of living and health care; (2) the stream of costs begins and the patient becomes completely disabled at the time of diagnosis; (3) the patient's annual costs are constant from diagnosis to death; and (4) the stream of the patient's annual costs is discounted to 1991 at a rate of 4%. Finally, because multiple reports of many epidemiological and cost parameters for Alzheimer's disease exist, interval estimates of the disease population and costs were constructed. The point estimates given below are based on midrange estimates of the individual parameters. The details of the estimation procedures are not presented in full but are available from the authors.

Results

The Epidemiology of Alzheimer's Disease

We estimated the prevalence rates of Alzheimer's disease by age and sex for persons 65 and older from data in three US studies.[2–4] Applied to the middle series US population forecast for 1991,[5] our estimates give a prevalence rate for Alzheimer's disease of 0.042 and a total of 1.35 million cases in the population 65 and older. Other reported rates are much higher,[6,7] possibly owing to diagnostic inaccuracies or the inclusion of mild cases.[8,9] The absence of data on the occurrence of this disease in persons under age 65 forced us to estimate the prevalence in this age group. We assumed that the prevalence rate for persons under age 45 is zero and that the fraction of Alzheimer's patients diagnosed before age 65 is the same as that with symptom onset before age 60 (diagnosis lags symptom onset by 3.5 to 5.5 years[10–15]). Based on the age distributions of Alzheimer's patients given in two US reports,[16,17] we estimated the number of patients aged 45 to 64 at 245 000 persons.

We found only one useful US study giving incidence rates of Alzheimer's disease by age and sex,[18] so we added data from Swedish and Israeli studies[19,20] to those results. Our estimated incidence rates yielded 266 000 newly diagnosed US cases of the disease in 1991.

The duration of survival with Alzheimer's disease from symptom onset falls with the patient's age,[16,21] but the effect of age on survival after diagnosis has not been studied. Hence, we assumed that survival after diagnosis is independent of

The authors are with the Department of Pharmaceutical Economics and Policy, University of Southern California, Los Angeles.

Requests for reprints should be sent to Joel W. Hay, PhD, Department of Pharmaceutical Economics and Policy, University of Southern California, 1985 Zonal Ave, Los Angeles, CA 90033.

This paper was accepted February 15, 1994.

Note. The views expressed here are solely those of the authors.

TABLE 1—Estimated Net Costs of Alzheimer's Disease per Person in 1991 (Midrange Estimates)

Annual (undiscounted) direct costs	
Diagnosis (first year only)	$ 1 450
Nursing home	7 570
Long-term mental hospital	392
Paid home care	3 140
Regular physician care	233
Acute care hospitalization	1 202
Other patient direct costs	0
Caregiver medical care	153
Total direct cost (first year only)	14 140
Total direct cost (second and later years)	12 690
Annual (undiscounted) indirect costs of unpaid home care	20 900
Total cost first year, excluding morbidity and mortality	35 040
Total cost second and later years, excluding morbidity and mortality	33 590
Total discounted direct cost[a]	47 581
Total discounted (direct cost + unpaid caregiver cost)[a]	123 556
Total discounted (direct cost + unpaid caregiver cost + disability and premature mortality cost)[a]	173 932

[a]Assuming a 4% annual discount rate and survival of 3.3 years for men, 4.3 years for women.

age. We assumed further that women live 4.3 years after diagnosis and men live 3.3 years—the averages of disease durations reported in two studies of Alzheimer's.[14,22]

The Costs of Alzheimer's Disease

Alzheimer's patients are cared for in the community, in nursing homes, and in mental hospitals, but the number of patients treated at each of these locations is unknown. We therefore estimated the number of patients living at each location before estimating the costs of care.

Nursing home costs. Extrapolating from data in the National Nursing Home Survey of 1985,[23] we estimated the average daily nursing home census at 1.743 million persons in 1991. Unfortunately, data on the percentage of residents with Alzheimer's disease vary widely. Estimates from the Nursing Home Survey vary from 6.3% to 9.4%,[23,24] but a national survey of Medicaid nursing home residents found that 23.5% were diagnosed with the disease.[25] The Nursing Home Survey data themselves indicated that 43.3% of all residents in 1985 were diagnosed with organic brain syndrome, and, assuming that 60% to 70% of organic brain syndrome is due to Alzheimer's disease,[26-29] this estimate implies a range of 26.0% to 30.3% of all residents with Alzheimer's. Accordingly, we determined the percentage of Alzheimer's patients in nursing homes in 1991 to be 10% to 30%, the midrange of which gave an estimate of 349 000 residents with the disease.

Based on our estimate of nursing home average daily census, national nursing home spending in 1991[30] amounted to $34 400 per annual resident. Since it is claimed that the cost for a demented patient is 12% to 20% higher than the average,[31,32] we put the gross annual nursing home cost for an Alzheimer's patient at $38 500 to $41 300. We estimated normal annual usage of nursing homes by persons aged 45 and older (the ratio of residents without Alzheimer's disease to the population at risk) at 0.016 to 0.020 years, and calculated the normal annual cost of food and shelter to be $4700 per person in 1991.[33] Hence, the estimated annual net cost of Alzheimer's nursing home care was $7570 per patient overall.

Mental hospital costs. We estimated the number of mental hospital patients in 1991 (185 000 persons) by projecting an annual time series of average daily censuses.[34] The number of Alzheimer's patients living in mental hospitals was estimated from data in two studies on residents with organic brain syndrome or probable Alzheimer's disease,[34,35] and was reckoned at 4000 to 7000 in 1991. The estimated gross annual cost of mental hospitalization was $113 600 per patient with Alzheimer's—obtained by inflating the overall average cost per patient in 1988[34] to 1991 levels by the medical care component of the consumer price index. Normal mental hospital usage was determined to be negligible so that, apart from normal living costs, the estimate of annual mental hospital costs was $392 per person with Alzheimer's disease.

The cost of community care. Community care refers both to nonmedical services provided to patients with Alzheimer's disease and to respite services for caregivers. The number of Alzheimer's patients living in the community (the Alzheimer's population less the number of institutionalized patients) was estimated at 1.240 million persons. Based on results from nine studies of community-dwelling demented or mentally impaired patients, we estimated the weekly number of paid and unpaid caregiver hours at 7.9 and 52.5, respectively.[31,36-44] Paid caregiver time was priced at $9.38 (the rate for home health care workers in 1991[45]), but that figure is at least 10% lower than the hourly cost (expressed in 1991 dollars) reported in three studies of caregiving for demented persons,[31,44,46] and it also excludes home care agency costs. We therefore set the hourly rate at $9.40 to $10.30 and assumed that the opportunity cost of unpaid caregiver time was the same. The resulting estimates of gross annual paid and unpaid caregiver time were $4050 and $26 900 for the home-dwelling patient. Because the normal amount of in-home caregiver time is negligible,[46-50] the annual net costs of paid and unpaid community care were $3150 and $20 900 per Alzheimer's patient overall.

Physicians' services: diagnosis. Diagnostic costs were estimated at $1450 per patient in 1991. The figure is our 1983 estimate (based on 1987 technology) of the cost inflated by the medical services component of the consumer price index.

Physicians' services: regular care. In five studies it has been reported that (mostly) elderly patients with Alzheimer's disease or dementia use 0.8 to 2.4 times as many physician visits as the average for all persons 65 and older (6 to 19.5 visits per year[36,42-44,48] vs 8.2 visits[51]). If the same multiple holds regardless of age, then patients with Alzheimer's disease average from 1.3 fewer to 9.2 more visits annually than all persons aged 45 and older (6.6 in 1990[51]). To cost these visits, we drew on 1990 physician survey data from the American Medical Association[52,53] inflated to 1991 price levels, and we estimated physicians' average revenue per visit at $59 in 1991. The (midrange) estimated annual net cost of outpatient physician services was therefore $233 per Alzheimer's patient in 1991.

Hospitalization. In various studies, persons with Alzheimer's disease or other mental impairment used the hospital 30% to 200% more than normal controls or the national average,[36,42,43,54] roughly the same as the normal population,[31,55] or less than normal persons.[56] Given these contradictory findings, we assumed that the average Alzheimer's patient used 0 to 200% more hospital days annually than a normal person aged 45 or older (1.37 days in 1990[51]). This would have meant an additional 0 to 2.7 inpatient days priced (apart from normal daily living costs) at $831 each,[57] plus one physician visit per day costing $59. The total net cost of hospital use was $1202 per patient in 1991.

Other direct costs. Other costs of Alzheimer's disease mentioned in the literature are associated with prescribed drugs, transportation, and health care visits to (or by) nonphysician providers,[31,36,42,43,58] but owing to scanty or contradictory evidence, we assumed that these costs were zero.

Caregiver costs. Caregiving for Alzheimer's disease causes economic costs only when it leads to an abnormal use or loss of economic resources. The costs of respite services and caregivers' loss of productive time have been counted in the costs of home care, but another cost occurs if the caregiver uses an abnormal amount of medical care. According to reports, caregivers of demented persons use 46% more physician visits and 71% more prescribed drugs than control subjects, and are more likely to be hospitalized as well[59,60]; however, it has also been found that physician usage and sick days are essentially the same for Alzheimer's caregivers and matched control subjects.[38] Hence, we estimated additional physician costs at 0 to $224, drug costs at 0 to $170 per caregiver-year (the upper bound is 0.71 times national per capita drug purchases in 1991[43]), and hospital costs at zero. Estimated net caregiver costs were $153 per Alzheimer's patient overall.

Loss of lifetime earnings owing to disability and premature mortality. The loss of the Alzheimer's patient's labor income dates from diagnosis and covers the period of a normal life expectancy.[61] We estimated lifetime earnings from an age-income profile provided to us by Dorothy Rice and her associates. The income streams were discounted at 4% and aggregated over all persons diagnosed with the disease in 1991. The earnings loss was estimated at about $50 380 per person.

Total costs. The midrange estimates of annual costs per Alzheimer's patient and costs per case are summarized in Table 1. The annual prevalence cost of the disease is the size of the Alzheimer's population times the average annual cost per patient. At midrange values of the parameters, estimated direct prevalence costs were $20.6 billion in 1991. Total annual unpaid caregiver costs were $33.3 billion, and the present value of the costs of disability and premature mortality were $13.4 billion for persons first diagnosed in 1991. Altogether, the 1991 annual prevalence costs of Alzheimer's disease were $67.3 billion. Supposing conservatively that the Alzheimer's population is constant over time, the present values of the annual direct and total costs of all future generations of Alzheimer's patients beginning in 1991 are about $536 billion and $1.75 trillion, respectively. These figures are the estimated discounted direct and total present values of resources that society will consume or lose because of this disease beginning in 1991.

Conclusion

The costs of caring for a patient with Alzheimer's disease, as reported here, are higher than those (for 1983) given in our earlier paper owing partly to price inflation and partly to new assumptions drawn from the literature. Adjusted to 1990/91 price levels, other (data-based) estimates of the annual direct and unpaid caregiver costs of Alzheimer's disease or mental impairment range from about $12 000 to $47 000 per patient.[31,32,36,41] Our figures ($33 600 to $35 000) fall toward the middle of this range, but comparisons between them and the other estimates are tenuous because of differences in methodology. Nevertheless, all the estimates indicate that the major economic burden of Alzheimer's is the cost of long-term home and institutional care. For home-dwelling patients, most of this burden—nearly $31 000 per person in 1991—is borne indirectly by caregivers, but much of the cost of institutionalization falls on the public at large. Assuming that 20% of all nursing home residents had Alzheimers, Medicaid spent $5.7 billion[30] in 1991 on these patients. To add some perspective on this cost, we estimate that the total medical cost of acquired immunodeficiency syndrome (AIDS) was no more than $4.2 billion in 1991 (assuming 77 000 cases[62] and a cost per case of $50 000 to $55 000[63]). Thus, even the Medicaid cost of Alzheimer's probably exceeds the total medical cost of a more publicized disease like AIDS. Clearly, the discovery of palliatives or cures for Alzheimer's disease would greatly benefit society as a whole. □

References

1. Hay J, Ernst R. The economic costs of Alzheimer's disease. *Am J Public Health.* 1987;77:1169–1175.
2. Bachman DL, Wolf PA, Linn R, et al. Prevalence of dementia and probable senile dementia of the Alzheimer type in the Framingham study. *Neurology.* 1992;42: 115–119.
3. Beard CM, Kokmen E, Offord K, et al. Is the prevalence of dementia changing? *Neurology.* 1991;41:1911–1914.
4. Pfeffer RI, Afifi AA, Chance JM. Prevalence of Alzheimer's disease in a retirement community. *Am J Epidemiol.* 1987;125: 420–435.
5. Evans DA, Funkenstein HH, Albert MS, et al. Prevalence of Alzheimer's disease in a community of older persons: higher than previously reported. *JAMA.* 1989;262:2551–2556.
6. Arenberg D. Misclassification of "probable senile dementia-Alzheimer's type" in the Baltimore longitudinal study of aging. *J Clin Epidemiol.* 1990;43:105–107.
7. Sapir M. Re: "prevalence of Alzheimer's disease in a retirement community." *Am J Epidemiol.* 1988;127:1093–1094.
8. Spencer G. Projections of the population of the United States, by age, sex, and race: 1988 to 2080. *Current Population Reports.* Series P-25, no. 1018. Washington, DC: US Bureau of the Census; 1989.
9. Sayetta RB. Rates of senile dementia—Alzheimer's type in the Baltimore longitudinal study. *J Chronic Dis.* 1986;39:271–286.
10. Mayeux R, Stern Y, Spanton S. Heterogeneity in dementia of the Alzheimer type: evidence of subgroups. *Neurology.* 1985;35: 453–461.
11. Heyman A, Wilkinson WE, Hurwitz BJ, et al. Early onset Alzheimer's disease: clinical predictors of institutionalization and death. *Neurology.* 1987;37:980–984.
12. Huff FJ, Growden JH, Corkin S, et al. Age at onset and rate of progression of Alzheimer's disease. *J Am Geriatr Soc.* 1987;30:27–30.
13. Knopman DS, Kitto J, Deinard S, et al. Longitudinal study of death and institutionalization in patients with primary degenerative dementia. *J Am Geriatr Soc.* 1988;36: 108–112.
14. Hier DB, Warach JD, Gorelick PB, et al. Predictors of survival in clinically diagnosed Alzheimer's disease and multi-infarct dementia. *Arch Neurol.* 1989;46: 1213–1216.
15. Walsh JS, Welch G, Larson EB. Survival of outpatients with Alzheimer-type dementia. *Ann Intern Med.* 1990;113:429–434.
16. Heston LL, Mastri AR, Anderson VE, et al. Dementia of the Alzheimer type: clinical genetics, natural history and associated conditions. *Arch Gen Psychiatry.* 1981;38: 1085–1090.
17. Centers for Disease Control. Alzheimer disease—California, 1985–87. *MMWR.* 1990;39:105–107.
18. Schoenberg BS, Kokmen E, Okazaki H. Alzheimer's disease and other dementing illnesses in a defined United States popula-

tion: incidence rates and clinical features. *Ann Neurol.* 1987;22:724–729.
19. Rocca WA, Amaducci LA, Schoenberg BS. Epidemiology of clinically diagnosed Alzheimer's disease. *Ann Neurol.* 1986;19: 415–424.
20. Hagnell O, Ojesjo L, Rorsman B. Incidence of dementia in the Lundby study. *Neuroepidemiology.* 1992;11:61–66.
21. Diesfeldt HFA, van Houte LR, Moerkens RM. Duration of survival in senile dementia. *Acta Psychiatr Scand.* 1986;73:366–371.
22. Barclay LL, Zemcov A, Blass JP, McDowell FH. Factors associated with the duration of survival in Alzheimer's disease. *Biol Psychiatry.* 1985;20:86–93.
23. The National Nursing Home Survey: 1985. Summary for the United States. Hyattsville, Md: US Dept of Health and Human Services; 1989.
24. Strahan G, Burns BJ. Mental Illness in Nursing Homes: United States, 1985. Hyattsville, Md: US Dept of Health and Human Services; 1991.
25. Manton KG, Vertrees JC, Woodbury MA. Functionally and medically defined subgroups of nursing home populations. *Health Care Financ Rev.* 1990;12:47–62.
26. Blass JP. Alzheimer's disease. *DM.* 1985;31: 1–69.
27. Gruenberg EM. Epidemiology of senile dementia. *Adv Neurol.* 1978;19:437–455.
28. Maleta GJ, Pirozzolo FJ, Thompson G, et al. Organic mental disorders in a geriatric outpatient population. *Am J Psychiatry.* 1982;139:521–523.
29. Mortimer JA. Epidemiological aspects of Alzheimer's disease. In: Pirozzolo FJ, Maleta GJ, eds. Advances in Neurogerontology. New York, NY: Praeger; 1980:307–332.
30. Letsch SW, Lazenby HC, Levit KR, et al. National Health Expenditures, 1991. *Health Care Financ Rev.* 1992;14:1–30.
31. Coughlin TA, Liu K. Health care costs of older persons with cognitive impairments. *Gerontologist.* 1989;29:173–182.
32. Huang L, Cartwright WS, Hu T. The economic cost of senile dementia in the United States, 1985. *Public Health Rep.* 1988;103:3–7.
33. Jacobs E, Shipp S. How family spending has changed in the US. *Month Labor Rev.* 1990;113:20–27.
34. Redick RW, Witkin MJ, Atay JE, et al. Specialty mental health system characteristics. In: Manderscheid RW, Sonnenschein MA, eds. *Mental Health, United States 1992.* Washington, DC: US Dept of Health and Human Services; 1992:1–141. DHHS publication SMA 92–1942.
35. Moak GS, Fisher WH. Alzheimer's disease and related disorders in state mental hospitals: data from a nationwide survey. *Gerontologist.* 1990;30:798–802.
36. Hu T, Huang L, Cartwright WS. Evaluation of the costs of caring for the senile demented elderly: a pilot study. *Gerontologist.* 1986;26:158–163.
37. Birkel RC. Toward a social ecology of the home-care household. *Psychol Aging.* 1987; 2:294–301.
38. Kiecolt-Glaser J, Glaser R, Shuttleworth EC, et al. Chronic stress and immunity in family caregivers of Alzheimer's disease victims. *Psychosom Med.* 1987;49:523–535.
39. Dura JR, Haywood-Niler E, Kiecolt-Glaser JK. Spousal caregivers of persons with Alzheimer's and Parkinson's disease dementia: a preliminary comparison. *Gerontologist.* 1990;30:332–336.
40. Cattanach L, Tebes JK. The nature of elder impairment and its impact on family caregivers' health and psychosocial functioning. *Gerontologist.* 1991;31:246–255.
41. Rice DP, Fox PJ, Max W, et al. The burden of caring for Alzheimer's disease patients. In: *Proceedings of the 1991 Public Health Conference on Records and Statistics.* Washington, DC: US Dept of Health and Human Services; 1992:119–124.
42. Lawton MP, Brody EM, Saperstein AR. A controlled study of respite service for caregivers of Alzheimer's patients. *Gerontologist.* 1989;29:8–16.
43. Lawton MP, Brody EM, Saperstein AR, et al. Respite services for caregivers: research findings for service planning. *Home Health Care Serv Q.* 1989;10:5–31.
44. Weinberger M, Gold DT, Divine GW, et al. Expenditures in caring for patients with dementia who live at home. *Am J Public Health.* 1993;83:338–341.
45. US Bureau of Labor Statistics. *Employment Earnings.* 1992;39:137.
46. Ruchlin HS, Morris JN, Gutkin CE, et al. Expenditures for long-term care services by community elders. *Health Care Financ Rev.* 1989;10:55–65.
47. Berk ML, Bernstein A. Use of home health services: some findings from the National Medical Care Expenditure Survey. *Home Health Care Serv Q.* 1985;6:13–23.
48. Mohide EA, Pringle DM, Streiner DL, et al. A randomized trial of family caregiver support in the home management of dementia. *J Am Geriatr Soc.* 1990;38:446–454.
49. Scharlach AE, Boyd SL. Caregiving and employment: results of an employee survey. *Gerontologist.* 1989;29:382–387.
50. Stoller EP. Formal services and informal helping: the myth of service substitution. *J Appl Gerontol.* 1989;8:37–52.
51. Adams PF, Benson V. Current Estimates from the National Health Interview Survey, 1990. *Vital Health Stat [10].* 1991;181.
52. Gonzalez ML, ed. *Socioeconomic Characteristics of Medical Practice 1990/1991.* Chicago, Ill: American Medical Association; 1991.
53. Gonzalez ML, ed. Socioeconomic Characteristics of Medical Practice 1992. Chicago, Ill: American Medical Association; 1992.
54. Weiler PG, Lubben JE, Chi I. Cognitive impairment and hospital use. *Am J Public Health.* 1991;81:1153–1157.
55. Colerick EJ, George LK. Predictors of institutionalization among caregivers of patients with Alzheimer's disease. *J Am Geriatr Soc.* 1986;34:493–498.
56. Welch HG, Walsh JS, Larson EB. The cost of institutional care in Alzheimer's disease: nursing home and hospital use in a prospective cohort. *J Am Geriatr Soc.* 1992;40:221–224.
57. Donham CS, Maple BT, Levit KR. Health care indicators for the United States. *Health Care Financ Rev.* 1992;3:173–179.
58. Wolf-Klein GP, Silverstone FA, Brod MS, et al. Are Alzheimer patients healthier? *J Am Geriatr Soc.* 1988;36:219–224.
59. George LK, Gwyther LP. Caregiver well-being: a multidimensional examination of family caregivers of demented adults. *Gerontologist.* 1986;26:253–259.
60. Haley WE, Levine EG, Brown SL, et al. Psychological, social, and health consequences of caring for a relative with senile dementia. *J Am Geriatr Soc.* 1987;35:405–411.
61. National Center for Health Statistics. *Month Vital Stat Rep.* 1992;40:16.
62. Centers for Disease Control and Prevention. *HIV/AIDS Surveill Year-End Edition.* 1993;5:17.
63. Hay JW, Kizer KW. Medi-Cal expenditures for persons with AIDS. *AIDS Public Pol J.* 1993;8:91–102.

The Burden and Outcomes Associated with Dehydration among US Elderly, 1991

Joan L. Warren, PhD, W. Edward Bacon, PhD, MPH, Tamara Harris, MD, MS, A. Marshall McBean, MD, MSc, Daniel J. Foley, MS, and Caroline Phillips, MS

ABSTRACT

Objectives. Dehydration has been underappreciated as a cause of hospitalization and increased hospital-associated mortality in older people. This study used national data to analyze the burden and outcomes following hospitalizations with dehydration in the elderly.

Methods. Data from 1991 Medicare files were used to calculate rates of hospitalization with dehydration, to examine demographic characteristics and concomitant diagnoses associated with dehydration, and to analyze the contribution of dehydration to mortality.

Results. In 1991, 6.7% (731 695) of Medicare hospitalizations had dehydration listed as one of the five reported diagnoses, a rate of 236.2/10 000 elderly Medicare beneficiaries. In 1991, Medicare reimbursed over $446 million for hospitalizations with dehydration as the principal diagnosis. Older people, men, and Blacks had elevated risks for hospitalization with dehydration. Acute infections, such as pneumonia and urinary tract infections, were frequent concomitant diagnoses. About 50% of elderly Medicare beneficiaries hospitalized with dehydration died within a year of admission.

Conclusions. Hospitalization of elderly people with dehydration is a serious and costly medical problem. Attention should be focused on understanding predisposing factors and devising strategies for prevention. (*Am J Public Health.* 1994;84:1265–1269)

Introduction

Dehydration is a common condition in the elderly, and, in 1991, dehydration was 1 of the 10 most frequent diagnoses reported for Medicare hospitalizations (Health Care Financing Administration [HCFA], unpublished data). Dehydration has been reported previously as the reason for hospitalization in 0.65% and 2.25% of all admissions to two hospitals.[1] In a case series of 15 187 hospital admissions in a single hospital, 1.1% of patients 60 years of age and older were found to have hypernatremia.[2] These studies, from a limited number of hospitals, highlight the lack of published information on hospitalizations with dehydration, including the incidence of dehydration in the elderly population, patterns of mortality, and concomitant health conditions. As the number of elderly people increases—especially the oldest old, who appear to be more prone to dehydration[3–5]—the importance of dehydration as a medical condition is likely to increase. Examination of national data on hospitalizations among the elderly can help delineate the extent of the burden resulting from dehydration and promote the need for strategies to prevent morbidity, mortality, and costs associated with this condition.

In this paper, we examine hospitalizations with dehydration among people 65 years of age and older. We first examine the burden of dehydration in the elderly by presenting the rates of hospitalizations with dehydration in 1991 as well as identifying specific demographic groups at risk of hospitalization for dehydration and at risk of dying within 30 days of admission. We then explore areas for potential interventions that might lessen the impact of dehydration. This analysis includes the identification of concomitant diagnoses that are reported frequently for hospitalizations with dehydration, which may improve understanding of the causes of this condition. In addition, we estimate the effect of dehydration on mortality by comparing, for 10 groups of principal diagnoses, early and late mortality following hospitalizations with and without dehydration as a concomitant diagnosis.

Methods

Data for this project were obtained from the Medicare Provider Analysis and Review (MEDPAR) files maintained by HCFA. These files, which are available from 1984 to the present, contain summarized records of all claims submitted by hospitals for services provided to Medicare beneficiaries. For each hospitalization, the MEDPAR record contains information about the beneficiary's age, race, and sex; up to five diagnoses classified according to the *International Classification of Diseases, 9th Revision, Clinical*

Joan L. Warren and A. Marshall McBean are with the Epidemiology Branch, Office of Research, Health Care Financing Administration, Baltimore, Md. W. Edward Bacon is with the National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Md. Tamara Harris, Daniel J. Foley, and Caroline Phillips are with the Epidemiology, Demography and Biometry Program, National Institute on Aging, Bethesda, Md.

Requests for reprints should be sent to Joan L. Warren, PhD, Epidemiology Branch, Office of Research, 2504 Oak Meadows Bldg, Health Care Financing Administration, 6325 Security Blvd, Baltimore, MD 21207.

This paper was accepted March 2, 1994.

Note. The views expressed here are the authors' and do not necessarily reflect those of the Health Care Financing Administration, the National Center for Health Statistics, or the National Institute on Aging.

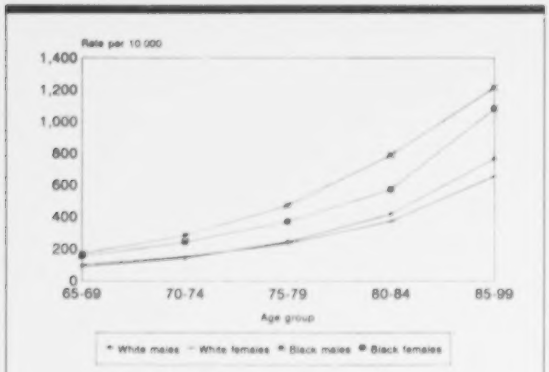


FIGURE 1—Rate of hospitalization with any listed diagnosis of dehydration per 10 000 elderly Medicare beneficiaries, 1991.

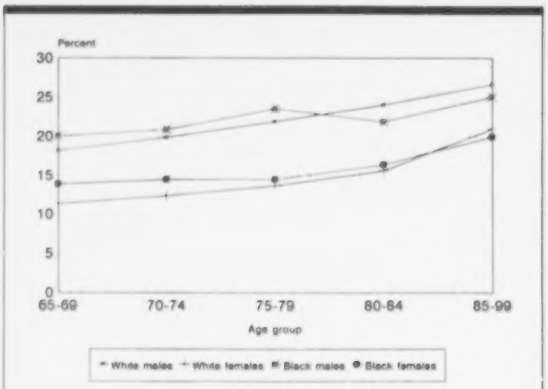


FIGURE 2—Age-, race-, and sex-specific percentages of patients dying within 30 days of admission, among elderly Medicare beneficiaries hospitalized with a diagnosis of dehydration, 1991.

Modification (ICD-9-CM)[6]; the amount reimbursed to the hospital by Medicare and the dates of admission and discharge; and the date of death (reported by the Social Security Administration) if death occurred within 2 years after the year of hospitalization.

Dehydration is a condition with multiple causes,[1,7-10] and the term is commonly used to reflect several related physiological states based on an imbalance between intake and loss of fluid and the accompanying sodium status.[4,10,11] The present study included as cases all hospitalizations involving an ICD-9-CM code for volume depletion (276.5), regardless of sodium status, since in most cases sodium status was not reported. Medicare beneficiaries who were hospitalized in 1991 with dehydration were identified from a review of a 20% random sample of all hospitalizations. All records with an ICD-9-CM diagnosis of volume depletion appearing as the principal or as a concomitant diagnosis were selected. (The principal diagnosis is the diagnosis that, after thorough review, is believed by the physician to be the cause for the hospitalization.) We use the phrase "hospitalization with dehydration" to refer to cases with dehydration as one of the five diagnoses; "hospitalization for dehydration" refers to those with dehydration as the principal diagnosis. We limited our analysis to Black and White Medicare beneficiaries who were 65 to 99 years of age and resided in 1 of the 50 states or the District of Columbia. We excluded the estimated 5% to 6% of Medicare beneficiaries who were enrolled in a health maintenance organization (HMO) at the time of hospitalization because Medicare utilization records for HMO enrollees are incomplete.

The rate of hospitalization with dehydration in 1991 was calculated per 10 000 elderly beneficiaries enrolled in Medicare's hospitalization insurance program (Part A) as of July 1 of that year. Preliminary data showed an interaction effect between increasing age and sex on rates of hospitalization with dehydration; therefore, we calculated age-, race-, and sex-specific rates. Comparisons between all possible combinations of strata were made (i.e., rates for age between each race–sex group, rates for sex between each race–age group, and rates for race between each sex–age group). The method described by Kahn and Sempos[12] was used to calculate the relative risk and surrounding 95% confidence interval for each comparison. The respective reference groups for each of the comparisons were individuals 65 to 69 years of age, men, and Whites. The proportion of people who died within 30 days of being admitted to the hospital with dehydration was calculated for each age, race, and sex group, and comparisons were made as described above.

To determine the cost to Medicare, we totaled the amount reimbursed in 1991 by HCFA for hospitalizations with dehydration as the principal diagnosis. The average cost per hospitalization was determined from the median amount reimbursed by HCFA per hospitalization for dehydration as the principal diagnosis. We were unable to determine from these data the additional costs to Medicare for cases involving dehydration as a concomitant diagnosis.

We used the MEDPAR data to examine all diagnoses reported for hospitalizations with dehydration in 1991. To facilitate interpretability of the data, diagnoses were assigned to 1 of 10 categories based on ICD-9-CM codes. (The specific codes included in each diagnosis category are listed in the Appendix.) The frequency at which each diagnosis category occurred for hospitalizations with dehydration is presented. Because each record contains up to five diagnoses, more than one diagnosis category may be present during a single hospitalization.

To examine the contribution of dehydration to mortality for selected principal diagnoses, death rates after hospitalization with a concomitant diagnosis of dehydration were compared with death rates after hospitalization with no concomitant diagnosis of dehydration. The 10 previously described diagnosis categories were used in classifying principal diagnoses. Hospitalizations without a concomitant diagnosis of dehydration were obtained from a 20% random sample of hospitalization records from 1991. For this comparison, the mortality rates were adjusted for race, sex, and age by the direct method, with the total number of beneficiaries hospitalized in 1991 as the standard population. Relative risks and surrounding 95% confidence intervals were calculated as described previously. Deaths occurring within 30 days of admission were considered as early mortality;

TABLE 1—Frequency of Selected Diagnoses Reported for Hospitalizations with Any Listed Diagnosis of Dehydration among Elderly Medicare Beneficiaries, 1991

Diagnosis Category[a]	Sample, %
Respiratory illness	28.2
Urinary system infections	24.9
Cardiac	21.8
Frailty	20.3
Other metabolic disorders	18.9
Other gastrointestinal conditions	18.8
Cancer	15.7
Diabetes	12.0
Gastroenteritis	10.4
Sepsis	7.1

[a]Diagnosis appears as one of five reported to Medicare; see Appendix for specific codes for each category.

TABLE 2—Early and Late Mortality Rates for Selected Principal Diagnosis Categories among Elderly Medicare Beneficiaries Hospitalized with and without a Concomitant Diagnosis of Dehydration, 1991

Principal Diagnosis Category	30-Day Post-admission Mortality Rate[a]	Relative Risk (95% Confidence Interval)	31- to 365-Day Post-admission Mortality Rate[a]	Relative Risk (95% Confidence Interval)
Dehydration	17.4	... ...	30.6	... ...
Respiratory illness				
Without dehydration[b]	10.9	1.00 ...	17.7	1.00 ...
With dehydration	18.9	1.73 (1.66, 1.79)	23.9	1.35 (1.31, 1.39)
Gastroenteritis				
Without dehydration[b]	3.1	1.00 ...	12.7	1.00 ...
With dehydration	3.3	1.07 (0.89, 1.30)	14.7	1.16 (1.06, 1.26)
Other gastrointestinal conditions				
Without dehydration[b]	4.9	1.00 ...	10.2	1.00 ...
With dehydration	8.8	1.79 (1.68, 1.91)	18.1	1.78 (1.70, 1.86)
Urinary system infections				
Without dehydration[b]	7.2	1.00 ...	18.2	1.00 ...
With dehydration	11.5	1.58 (1.47, 1.70)	29.6	1.62 (1.56, 1.69)
Cancer				
Without dehydration[b]	14.1	1.00 ...	25.9	1.00 ...
With dehydration	45.2	3.20 (3.10, 3.30)	40.4	1.56 (1.50, 1.62)
Sepsis				
Without dehydration[b]	22.5	1.00 ...	20.9	1.00 ...
With dehydration	29.5	1.31 (1.25, 1.37)	27.2	1.30 (1.23, 1.38)
Cardiac				
Without dehydration[b]	10.4	1.00 ...	14.9	1.00 ...
With dehydration	22.8	2.19 (2.07, 2.31)	26.5	1.78 (1.69, 1.88)
Frailty				
Without dehydration[b]	16.9	1.00 ...	25.4	1.00 ...
With dehydration	25.5	1.51 (1.43, 1.60)	36.1	1.42 (1.35, 1.49)
Diabetes				
Without dehydration[b]	5.1	1.00 ...	13.7	1.00 ...
With dehydration	8.9	1.73 (1.50, 1.99)	18.1	1.32 (1.20, 1.45)
Other metabolic disorders				
Without dehydration[b]	7.9	1.00 ...	17.3	1.00 ...
With dehydration	11.1	1.41 (1.20, 1.65)	24.4	1.41 (1.27, 1.57)

[a]Per 100 hospitalizations. Adjusted for age, race, and sex.
[b]Reference group.

those occurring between 31 and 365 days after admission were considered as late mortality.

Results

In 1991, the rate of hospitalization per 10 000 elderly Medicare beneficiaries was 236.2 admissions with any listed diagnosis of dehydration and 49.7 admissions for dehydration as a principal diagnosis. Of the more than 10 million hospitalizations involving elderly Medicare beneficiaries during that year, 6.7% (731 695 hospitalizations) had a diagnosis of dehydration in any position and 1.4% (146 960 hospitalizations) had dehydration listed as the principal diagnosis.

There were significant differences in the rate of hospitalization with dehydration by age, race, and sex (Figure 1). Increasing age was a major factor in risk of admission with dehydration, with people 85 to 99 years of age more than 6 times more likely to be hospitalized for dehydration than those 65 to 69 years of age. At each age, Blacks were 1.5 to 2 times more likely than Whites to be hospitalized with dehydration. Within race, men were more likely to be admitted with dehydration than were women, except for Whites 65 to 79 years of age.

Overall, more than 18% of patients died within 30 days of hospitalization with dehydration. Mortality rates increased with age, even among the young old (those 65 to 74 years of age), and 30-day postadmission mortality was about 12% for women and 18% for men. For each age–race stratum, the proportion of men dying was 26% to 63% higher than that of women (Figure 2). Comparison of data by race suggests that, for both sexes, the percentage of Blacks dying was higher up to 79 years of age but lower for the 85- to 99-year age range. The greatest difference between the race–sex groups was for Black women 65 to 69 years old, who had mortality rates 22% higher than White women in the same age group.

The total amount reimbursed by HCFA to hospitals for treatment of beneficiaries hospitalized for dehydration as the principal diagnosis was more than $446 million. The median cost per hospitalization for dehydration as the principal diagnosis was $2942.

A review of all diagnoses reported for hospitalizations with dehydration showed that a large number of patients with dehydration also had acute infections: 28.2% had a respiratory illness, 24.9% had a diagnosis of urinary system infection, 10.4% had a diagnosis of gastroenteritis, and 7.1% had a diagnosis of sepsis (Table 1). Chronic conditions such as frailty, cancer, and diabetes occurred in 20.3%, 15.7%, and 12.0% of hospitalizations, respectively.

Among people hospitalized with a principal diagnosis of dehydration, almost half died within a year of admission, with 17.4% dying within 30 days of admission

and an additional 30.6% dying in the remaining 11 months of the year (Table 2). For hospitalizations with dehydration as a concomitant condition, there was considerable variation in mortality based on the principal diagnosis. Comparison of mortality rates by principal diagnosis for hospitalizations with and without a concomitant diagnosis of dehydration showed that, for each principal diagnosis, hospitalizations with a concomitant diagnosis of dehydration involved significantly higher mortality for both the early and later periods, with the exception of deaths within 30 days of hospitalization for gastroenteritis.

Discussion

These data demonstrate that dehydration in the elderly is a serious, frequent, and costly medical problem that may not have been fully recognized to date. Older people, men, and Blacks are more likely to be hospitalized for dehydration. Hospitalization resulting from dehydration is associated with an increased risk of death, particularly for older men. In 1991, Medicare payments were more than $446 million for beneficiaries hospitalized with dehydration as a principal diagnosis. This figure does not include costs of treatment for the entire episode of care (physician visits, home health or skilled nursing care), care provided to beneficiaries enrolled in an HMO, or additional costs to Medicare for 584 735 non-HMO hospitalizations that involved dehydration as a complicating condition.

In this study, the 1.4% of hospitalizations with dehydration as the principal diagnosis is comparable to what has been reported in previous research.[1,2] Unlike prior researchers, we reviewed all diagnoses reported for each hospitalization. We found that 6.7% of all hospitalizations had dehydration listed as one of five diagnoses. This finding suggests that all diagnoses reported during a hospitalization should be reviewed to assess the full extent of dehydration as a problem among hospitalized elderly people.

In the present study, of people admitted with a principal diagnosis of dehydration, the percentage dying within 30 days of admission—17.42—was lower than the 46% to 48% reported for hospitalizations for hypernatremic dehydration.[3,13] The variation in mortality found between the present and previous studies may reflect differences in case definition or special characteristics of those with hypernatremia. However, previous studies have attributed high mortality rates not to the level of hypernatremia but to the significant underlying disease.[3,13] To address this issue, we stratified by principal diagnosis. From this stratification, it appears that the risk of mortality after hospitalization with dehydration varies considerably depending on principal diagnosis. However, for each principal diagnosis, the concomitant diagnosis of dehydration was associated with an increased risk of death.

The high rates of mortality suggest that hospitalization with dehydration should be considered a marker for an increased risk of death within the next year. We examined another common condition in the elderly, hip fracture, that has a high post-admission mortality rate and is often considered to place the person at increased risk of death within the year following the event. By comparison, mortality within 30 days of admission has been reported as 6.4% of hospitalizations with hip fracture[14] and 17.4% of hospitalizations with dehydration. Within 1 year, 20% of patients with hip fractures died,[15] in comparison with more than 47% of those with dehydration.

Recognizing the incidence, risks, and costs associated with dehydration raises the need for strategies to prevent the condition. Of the underlying diseases likely to cause dehydration, some appear preventable. Respiratory illnesses, which were diagnosed in more than 28% of those hospitalized with dehydration, could possibly be prevented or reduced by influenza and pneumococcal vaccination.[16,17] Yet, it is currently estimated that only 30% to 41% of individuals over age 65 are immunized annually against influenza, and only 14% to 20% have received a pneumococcal vaccination.[18,19] Other infectious diseases, although not as amenable to prevention, should be recognized as potential predisposing factors for dehydration among older people. For instance, dehydration associated with gastroenteritis may be prevented by prompt recognition of the gastroenteritis and intervention through fluid orders, including the use of oral rehydration or administration of intravenous fluids. Efforts to prevent dehydration should be directed toward educating elderly patients, especially those living alone, regarding the risk of dehydration and the need to maintain adequate oral intake during periods of acute illness. Nursing personnel should also be made aware of the importance of monitoring fluid intake in elderly patients in hospitals and nursing homes.

There are several limitations to our study. We cannot determine from the data whether the dehydration occurred prior to or during the hospitalization. In addition, we cannot ascertain from the Medicare data the clinical treatment that might have influenced dehydration outcomes (e.g., rapid rates of rehydration that might increase mortality).[3] Administrative data do not provide information on the underlying health status of people hospitalized with dehydration. While it appears that acute illness is associated with dehydration, it may be that these individuals also have chronic medical conditions that influence the risk and outcomes of dehydration.

The accuracy of secondary data in measuring the extent of dehydration may be questioned. However, the MEDPAR data used in the present study were validated from a second source, the National Hospital Discharge Survey. This survey, conducted annually by the National Center for Health Statistics (NCHS), includes data abstracted from the medical records of a representative sample of discharges from non-federal short-stay hospitals in the United States.[20] The survey data are subject to field quality checks for validity and reliability. For the years 1984 to 1991, a comparison of the data available from the survey and from MEDPAR revealed that the population-based rates of dehydration observed from both sources were essentially the same (HCFA and NCHS, unpublished data, 1984–1991).

In conclusion, this study provides the first national estimates that demonstrate the extent to which dehydration is a significant cause of morbidity, mortality, and health care costs in the elderly. Hospitalizations with dehydration are associated with acute illness, which suggests that some cases of dehydration may be preventable. Investigations in the clinical setting of specific diseases associated with dehydration may yield additional preventive strategies and sources of cost savings. Furthermore, rates of mortality after hospitalizations with dehydration both as a principal and as a concomitant diagnosis are high. People hospitalized with dehydration should be considered at increased risk for dying within the year following admission. Recognition by health care providers of the extent that dehydration is a problem in the elderly, prevention of the underlying causes, and prompt treatment of the early stages of dehydration may reduce the morbidity, mortality, and costs resulting from this condition. □

References

1. Himmelstein DU, Jones AA, Woolhandler S. Hypernatremic dehydration in nursing home patients: an indicator of neglect. *J Am Geriatr Soc.* 1983;31:466–471.
2. Snyder NA, Fiegal DW, Arieff AI. Hypernatremia in elderly patients. *Ann Intern Med.* 1987;107:309–319.
3. Phillips PA, Rolls BJ, Ledingham JG, et al. Reduced thirst after water deprivation in healthy elderly men. *N Engl J Med.* 1984; 311:753–759.
4. Lavizzo-Mourey R, Johnson J, Stolley P. Risk factors for dehydration among elderly nursing home residents. *J Am Geriatr Soc.* 1988;36:213–218.
5. Lye M. Disturbances of homeostasis. In: Brocklehurst JC, Tallis RC, Fillit HM, eds. *Textbook of Geriatric Medicine and Gerontology.* New York, NY: Churchill Livingstone; 1992.
6. *International Classification of Diseases, 9th Revision, Clinical Modification.* Vol. 1. Washington, DC: US Dept of Health and Human Services; 1988.
7. Gault MH, Dixon ME, Doyle M, et al. Hypernatremia, azotemia, and dehydration due to high-protein tube feeding. *Ann Intern Med.* 1968;68:778–791.
8. Silver AJ. Aging and risks for dehydration. *Cleve Clin J Med.* 1990;57:341–344.
9. Lavizzo-Mourey RJ. Dehydration in the elderly: a short review. *J Natl Med Assoc.* 1987;79:1033–1038.
10. Alpern RJ, Saxton CR, Seldin DW. Diagnosis and management of fluid, electrolyte, and acid-base disorders. In: Kokko JP, Tannen RL, eds. *Fluids and Electrolytes.* Philadelphia, Pa: WB Saunders; 1990.
11. Leaf A. Dehydration in the elderly. *N Engl J Med.* 1984;311:791–792.
12. Kahn HA, Sempos CT. *Statistical Methods in Epidemiology.* New York, NY: Oxford Press; 1989.
13. Mahowald JM, Himmelstein DU. Hypernatremia in the elderly: relation to infection and mortality. *J Am Geriatr Soc.* 1981;29: 177–180.
14. *Health Care Financing Special Report: Hospital Data by Geographic Area for Aged Medicare Beneficiaries: Selected Procedures, 1986, Volume 2.* Baltimore, Md: Health Care Financing Administration; 1990. Publication 03301.
15. Magaziner J, Simonsick EM, Kashner TM, et al. Survival experience of aged hip fracture patients. *Am J Public Health.* 1989;79:274–278.
16. Shapiro ED, Clemens JD. A controlled evaluation of the protective efficacy of pneumococcal vaccine for patients at high risk of serious pneumococcal infections. *Ann Intern Med.* 1984;101:325–330.
17. Shapiro ED, Berg AT, Austrian R, et al. The protective efficacy of polyvalent pneumococcal polysaccharide vaccine. *N Engl J Med* 1991;325:1453–1460.
18. Gardener P, Schaffner W. Immunization of adults. *N Engl J Med.* 1993;328:1252–1258.
19. *Background for Report to Congress on the Medicare Influenza Vaccine Demonstration.* Baltimore, Md: Health Care Financing Administration; 1993.
20. Graves EJ. National Hospital Discharge Survey: annual summary, 1990. *Vital Health Stat [13].* 1990;112.

APPENDIX—ICD-9-CM Codes Used to Classify Diagnoses for Medicare Beneficiaries Hospitalized with Dehydration

Diagnosis Category	ICD-9-CM Code(s)
Dehydration	276.5
Respiratory illness	
Acute bronchitis	466.0–466.1
Chronic lung disease	490.0–496
Pneumonia and influenza	480–487.8
Gastroenteritis	003.0, 008.5, 008.6, 008.8, 009.0–009.3, 536.2, 558.9, 787.0
Other gastrointestinal conditions	
Abdominal pain	789.0
Cholelithiasis	574.0–574.51
Diseases of the esophagus	530.0–530.9
Diverticulitis	562.01, 562.03, 562.11, 562.13
Diverticulosis	562.00, 562.02, 562.10, 562.12
Gastritis	535.0–535.6
Gastrointestinal hemorrhage	578.0–578.9
Idiopathic proctocolitis	556
Intestinal obstruction	560.8–560.99
Pancreatitis	577.0–577.9
Paralytic ileus	560.1
Regional enteritis	555.0–555.9
Urinary system infection	
Kidney infection	590.0–590.9
Urinary tract infection	595.0, 595.9, 599.0
Cancer	140.0–208.9
Sepsis	038.0–038.9
Cardiac	
Acute myocardial infarction	410.0–410.92
Angina	411.1
Arrhythmias	427.0–427.32, 427.6–427.9
Congestive heart failure	428.0–428.9
Frailty	
Aspiration pneumonia	507.0
Decibitus ulcer	707.0
Dementia	290.0–290.9, 331.0, 331.2, 797
Impaction	560.3–560.39
Malnourishment	263.0–263.9, 262, 783.0, 799.4, 260, 261
Diabetes	250.0–250.91
Other metabolic disorders	
Diabetes insipidus	253.5, 588.1
Hyperkalemia	276.7
Hypernatremia	276.0
Hypoglycemia	251.2
Hypokalemia	276.8
Hyponatremia	276.1
Hypotension	458.0–458.9
Inappropriate antidiuretic hormone (SIADH)	253.6

The Groningen Activity Restriction Scale for Measuring Disability: Its Utility in International Comparisons

ABSTRACT

Objectives. The Groningen Activity Restriction Scale (GARS) is a non–disease-specific instrument to measure disability in activities of daily living (ADL) and instrumental activities of daily living (IADL). It was developed in studies of Dutch samples consisting of elderly or chronically ill people. The psychometric properties of the GARS demonstrated in these studies were highly satisfactory. This paper addresses the psychometric properties of the GARS across countries.

Methods. Data of 623 patients with recently diagnosed rheumatoid arthritis from four European countries were analyzed by means of a principal components analysis and a Mokken scale analysis for polychotomous items.

Results. The results of the analyses were highly satisfactory: there was one strong and reliable general factor representing one underlying dimension of disability in ADL and IADL, and there was a clear hierarchical ordering of the items included in the GARS. The validity of the GARS was strongly suggested by the pattern of associations of the GARS with age, sex, and other existing health status measures.

Conclusions. The psychometric characteristics of the GARS, which measures disability in ADL and IADL simultaneously, make this instrument very useful for comparative research across countries. (*Am J Public Health.* 1994;84:1270–1273)

Theo P. B. M. Suurmeijer, PhD, Dirk M. Doeglas, MA, Torbjorn Moum, PhD, Serge Briançon, MD, PhD, Boudien Krol, MA, Robbert Sanderman, PhD, Francis Guillemin, MD, Anders Bjelle, MD, PhD, and Wim J. A. van den Heuvel, PhD

Introduction

Rheumatoid arthritis is a chronic disease with a varying clinical course. Therefore, measures of the "objective" severity of the disease are always time-dependent. The paroxysmal character of the disease makes it fairly unpredictable for patients themselves and for their significant others. Part of the burden of rheumatoid arthritis is caused by this varying and unpredictable course.[1]

Sooner or later, patients with rheumatoid arthritis experience impairments and "social disability," that is, "dysfunctioning in a social role or in some aspect of social behavior."[2] The relevance of the behavioral consequences of a disease has often been stressed because these consequences are of crucial importance for the daily functioning of the individuals involved.[3] In this article we will confine ourselves to restrictions regarding certain activities and tasks, that is, activities of daily living (ADL) and instrumental activities of daily living (IADL), sometimes referred to as "housekeeping activities of daily living."[4]

ADL functions are essential for an individual's self-care (e.g., washing or dressing oneself), whereas IADL functions are more concerned with self-reliant functioning in a given environment (e.g., shopping, preparing meals).[5] The distinction between these two groups of activities is mainly a consequence of "institutional thinking." In many countries, the delivery of care with respect to these two groups of activities is provided by different professions or agencies (sometimes combined with different types of financing); therefore, these two groups of activities have been distinguished and measured separately.[6,7] From the perspective of the patient, however, ADL functions are no less "instrumental" than IADL functions.

Partly on the basis of existing instruments, Kempen and Suurmeijer developed the Groningen Activity Restriction Scale (GARS).[8,9] The instrument was developed to measure both ADL and IADL disability in community-based studies with respect to the aid and services provided by professional home help and district nursing agencies. The patient's physical condition, of course, plays an important role in this respect, but the patient's personality and contextual and interactional variables may be important as well in the "appraisal of one's abilities."[9,10]

The GARS has been applied in several studies in the Netherlands.[11] It has proven to be a very useful instrument: it makes it possible to (1) more precisely describe the severity of the disablement caused by several chronic conditions, (2) establish changes in disablement over

Theo P. B. M. Suurmeijer, Dirk M. Doeglas, Boudien Krol, Robbert Sanderman, and Wim J. A. van den Heuvel are with the Northern Centre for Health Care Research, University of Groningen, Groningen, the Netherlands. Dr Suurmeijer and Dr van den Heuvel are also with the Departments of Sociology and Health Sciences, University of Groningen. Torbjorn Moum is with the Department of Behavioral Sciences in Medicine, University of Oslo, Oslo, Norway. Serge Briançon and Francis Guillemin are with the Institute of Public Health, University of Nancy, Nancy, France. Anders Bjelle is with the Department of Rheumatology, Sahlgren University Hospital, Göteborg, Sweden.

Requests for reprints should be sent to Theo P. B. M. Suurmeijer, PhD, Northern Centre for Health Care Research, A. Deusinglaan 1, 9713 AV Groningen, the Netherlands.

This paper was accepted February 4, 1994.

time, (3) differentiate more accurately between degrees of disability, and (4) improve the assessment of the need for professional care.

In this article we use data from an international longitudinal study of patients recently diagnosed with rheumatoid arthritis (European Research on Incapacitating Diseases and Social Support [EURIDISS][12]) to analyze the overall psychometric properties of the GARS across countries. More specifically, we test the unidimensionality and hierarchical order of the scale items and explore the scale's construct validity.

Methods

Sampling Procedures

The criteria for inclusion in the EURIDISS study have been published elsewhere.[12] By the end of November 1992, data had been collected on a total of 630 patients with rheumatoid arthritis: 116 French, 292 Dutch, 124 Norwegian, and 98 Swedish patients. The nonresponse rate varied from 12% (the Netherlands) to 30% (France). The patients' mean age was 52 years; mean disease duration was 2.6 years. Thirty-one percent of the patients were men and 69% were women. Because of incomplete data, 7 of these respondents were omitted from the analysis.

Description of the Groningen Activity Restriction Scale (GARS)

The 18 GARS items and five response categories are presented in Table 1. The items refer to what respondents are *able* to do and not to their actual performance, which is a very important distinction.[13] When an item refers to more than one activity (e.g., item 5), the activity causing the greatest problems to the patient determines the response.

Statistical Analysis

In the analyses, the two most extreme responses (4 and 5) were combined, partly because response 5 was chosen by only a few patients. Consequently, sum scores could vary from 18 (not disabled) through 72 (severely disabled).

A principal components analysis was carried out on the scale items, both for the countries separately and for the total sample. Next, reliability figures (Cronbach's alpha[14] or rho[15]) were calculated.

The hierarchical order of the ADL and IADL items was tested with the Mokken scale analysis for polychotomous items (MSP).[15,16] MSP is a probabilistic (nonparametric) extension of Guttman scale analysis[17] that can handle three or more rank-ordered response categories per item. The program calculates three different scalability coefficients: H_i for individual items, H_{ij} for item pairs, and H for a set of items as a whole. For a set of items to be accepted as a scale, it is required that all H_i's be greater than .00 and all H_{ij}'s be greater than or equal to .30. In that case H will be greater than or equal to .30. The H coefficient refers to the strength of the scale as a whole: if H is from .30 up to .40, the scale is "weak"; if H is from .40 up to .50, the scale is "moderately strong"; if H is equal to or greater than .50, the scale is "strong."

Analyses of variance were carried out to test for differences in GARS scores across countries, age, and sex, and the construct validity of the GARS and its component parts was explored.

TABLE 1—Items and Response Categories of the Groningen Activity Restriction Scale (GARS)

Response categories for each item
1. Yes, I can do it fully independently without any difficulty.
2. Yes, I can do it fully independently but with some difficulty.
3. Yes, I can do it fully independently but with great difficulty.
4. No, I cannot do it fully independently, I can only do it with someone's help.
5. No, I cannot do it at all, I need complete help.

Activities of daily living (ADL)
1. Can you, fully independently, dress yourself?
2. Can you, fully independently, get in and out of bed?
3. Can you, fully independently, stand up from sitting in a chair?
4. Can you, fully independently, wash your face and hands?
5. Can you, fully independently, wash and dry your whole body?
6. Can you, fully independently, get on and off the toilet?
7. Can you, fully independently, feed yourself?
8. Can you, fully independently, get around in the house (if necessary, with a cane)?
9. Can you, fully independently, go up and down the stairs?
10. Can you, fully independently, walk outdoors (if necessary, with a cane)?
11. Can you, fully independently, take care of your feet and toenails?

Instrumental activities of daily living (IADL)
12. Can you, fully independently, prepare breakfast or lunch?
13. Can you, fully independently, prepare dinner?
14. Can you, fully independently, do "light" household activities (for example, dusting and tidying up)?
15. Can you, fully independently, do "heavy" household activities (for example, mopping, cleaning the windows, and vacuuming)?
16. Can you, fully independently, wash and iron your clothes?
17. Can you, fully independently, make the beds?
18. Can you, fully independently, do the shopping?

Results

Results of Factor Analyses

From the principal components analysis on the GARS items, both for the countries separately and for the total sample, it appeared that the first factor was a very strong and reliable one. The sum score of the GARS was directly derived from the raw (i.e., unweighted) scores. The correlation between this sum score and the factor scores (i.e., weighted item scores) was .9938. Therefore, the raw scores can be used when calculating the GARS score. A scree plot of the eigenvalues of the extracted components confirmed the assumption of one underlying dimension: after the first general component (eigenvalue = 8.71), there was a clear "elbow" in the curve of the consecutive eigenvalues.

As mentioned before, the ADL and IADL items are often considered to indicate two specific types of disability. Therefore, we repeated the same analyses as before for the 11 ADL and 7 IADL items separately, for the countries separately and for the total sample. It appeared that both the ADL and IADL items form one strong and reliable general factor; no other factors with an eigenvalue of 1.00 or more could be extracted.

The mean GARS score (range = 29.2–32.6), as well as the mean ADL and IADL scores (range = 15.8– 17.4 and 13.2–15.2, respectively), differed some-

TABLE 2—Correlations of the GARS and ADL and IADL Scales with the Physical Mobility Subscale from the Nottingham Health Profile (NHP-PM), the Karnofsky Physical Status Scale (KPSS), the Overall Evaluation of Health (OEH), and the Somatic Symptoms Subscale from the General Health Questionnaire (GHQ-SS)

	NHP-PM	KPSS	OEH	GHQ-SS
GARS	.78	.68	−.40	.25
ADL	.77	.64	−.39	.26
IADL	.71	.66	−.37	.21

Note. GARS = Groningen Activity Restriction Scale; ADL = activities of daily living; IADL = instrumental activities of daily living.

TABLE 3—Mokken Scale Analysis for Polychotomous Items on the Groningen Activity Restriction Scale (GARS), for the Combined European Data Set (n = 623)

Ordered GARS Items		Item Mean Scores	Scalability Coefficient for the Individual Items (H_i)
1. Wash face/hands	(ADL)	1.22	.49
2. Feed yourself	(ADL)	1.24	.40
3. Get around inside house	(ADL)	1.26	.50
4. Get on/off toilet	(ADL)	1.32	.48
5. Prepare breakfast/lunch	(IADL)	1.35	.47
6. Get in/out of bed	(ADL)	1.43	.47
7. Stand up from chair	(ADL)	1.45	.49
8. Do light cleaning	(IADL)	1.54	.46
9. Dress yourself	(ADL)	1.59	.53
10. Walk outdoors	(ADL)	1.60	.50
11. Wash/dry body	(ADL)	1.76	.51
12. Prepare dinner	(IADL)	1.79	.47
13. Go up/down stairs	(ADL)	1.81	.48
14. Wash/iron clothes	(IADL)	2.07	.47
15. Take care of feet/toenails	(ADL)	2.10	.42
16. Make beds	(IADL)	2.35	.47
17. Do shopping	(IADL)	2.49	.43
18. Do heavy cleaning	(IADL)	2.79	.49
Scalability coefficient *H* of the GARS			.47
Reliability coefficient rho			.94
Scalability coefficient *H* of the ADL items			.52
Reliability coefficient rho			.90
Scalability coefficient *H* of the IADL items			.51
Reliability coefficient rho			.89

what between countries. These differences may be due to environmental and cultural differences in the appraisal of activities included in the GARS. Probably because of sex role–specific socialization patterns, women scored significantly higher than men on the IADL scale, although the mean difference was very small (means = 14.8 and 13.6, respectively). On average, older people scored significantly higher on the GARS as a whole and on the ADL and IADL scales; these higher scores support the fact that, in general, older people are more disabled than younger ones[3] (range = 32.5–27.6, 17.5–14.8, and 15.0–12.8, respectively). There were no interaction effects between age, sex, and country.

The validity of the GARS and the ADL and IADL scales was explored by assessing the scale's association with several other instruments measuring physical problems and subjective health. Specifically, we used the Physical Mobility subscale from the Nottingham Health Profile,[18,19] the Karnofsky Performance Status Scale,[20] the Somatic Symptoms subscale from the General Health Questionnaire,[21] and an overall evaluation of health item using a ruler as a visual analogue scale. In the latter case, the patient was asked, "How would you rate your health at the moment? Would you say that it is very poor, that it is excellent, or that it is somewhere in between?"[22] The ruler ran from "very poor" (a score of 1) to "excellent" (a score of 100).

The GARS and the ADL and IADL scales were expected to correlate highest with those measures most comparable to the GARS (the Physical Mobility subscale from the Nottingham Health Profile and the Karnofsky Performance Status Scale) and to correlate lowest with those instruments less comparable to the GARS (the overall evaluation of health item and the Somatic Symptoms subscale from the General Health Questionnaire) (Table 2).

The pattern of correlations encountered yielded additional support for the utility of the GARS. Moreover, from the pattern and size of the correlation coefficients, it appears that there was no difference between the GARS and the ADL and IADL scales, which may be an additional reason to treat the ADL and IADL scales as one scale. (It also appears legitimate to use the ADL and IADL scales separately if necessary.)

Results of MSP

Finally, the hierarchical ordering of the items (in the data sets of the four European countries together) was tested with MSP (Table 3). The items are ordered according to their difficulty, as expressed by the item mean scores. All the H_i's were greater than or equal to .40; the reliability coefficient rho was .94. The H coefficient of .47 nearly met the criterion for a strong scale.

Conclusions

The results of the analyses were highly satisfactory. They showed a strong and reliable general first factor, indicating one underlying dimension of disability for ADL, IADL, or both. For institutional reasons, or to differentiate between sex or age groups, the ADL and IADL scales can be used separately.

The GARS turned out to be a rather strong unidimensional, hierarchical scale. The same was even more true for the separate ADL and IADL scales. The fact that the GARS items were hierarchically ordered means that respondents with the same score had the same problems with ADL and IADL. However, Kempen et al. warn against the use of cutpoints or single

questions to categorize patients: "Hierarchical evidence should not be used to justify minimal patient questioning."[23] This has partly to do with the fact that the Mokken model is not sample independent. Differences between samples may be due to language, education, motivation, opportunity, or environmental factors (living arrangements, availability and proximity of services, means of transportation, and so on).[10,13,23–25] However, when the Mokken analysis is applied to a maximally large group, its sample dependence gradually diminishes.

One of the advantages of MSP over the Guttman analysis is that more than two response categories can be used. As a consequence, more accurate distinctions can be made, which are not only more realistic about the daily functioning of patients but also necessary to target support services.[6,8,13,23]

Of course, other instruments have been developed to measure disability in ADL or IADL,[26,27] but the GARS measures both simultaneously, and it is a very reliable and valid scale with hierarchically ordered, polychotomous items. In addition, it is community based and not disease-specific. These characteristics make the GARS very useful for comparative and longitudinal research both across countries and across diseases. □

Acknowledgments

The EURIDISS study is supported in France by contract CRE 891102 from INSERM and a grant from the Société Française de Rhumatologie; in the Netherlands by Het Nationaal Reumafonds and by the Ministry of Welfare, Health and Cultural Affairs; in Norway by the Norwegian Research Council for Medicine; and in Sweden by the Riksförbundet mot Reumatism, Stockholm, and Göteborgs Läkaresällskap. Internationally, the study is supported by COMAC–Health Services Research contract MR4*0344-NL from the European Community.

The authors would like to thank rheumatologists S. van der Burg (Martini Hospital, Groningen), H. Lim (Scheper Hospital, Emmen), M. van Rijswijk and M. van Leeuwen (Academic Hospital, Groningen), F. Speerstra (Wilhelmina Hospital, Assen, and Refaja Hospital, Stadskanaal) (all from the Netherlands); A. Gaucher and J. Poure! (Department of Rheumatology, Nancy, France); and T. Kvien and L. M. Smedstad (Norwegian Lutheran Hospital, Oslo, Norway) for their collaboration. We would also like to thank G. I. J. M. Kempen for his very valuable remarks. In particular, we want to thank all the patients; without their cooperation this study would not have been possible.

A copy of a much more detailed version of this article is available from the first author on request.

References

1. Wiener CL. The burden of rheumatoid arthritis: tolerating the uncertainty. *Soc Sci Med.* 1975;9:97–104.
2. Wiersma D. Psychological impairments and social disabilities: on the applicability of the ICIDH to psychiatry. *Int Rehab Med.* 1986;8:3–7.
3. CBS/NIMAWO. *Lichamelijke beperkingen bij de Nederlandse bevolking.* 's-Gravenhage, the Netherlands: SDU-Uitgeverij; 1990.
4. Van den Bos GAM, Vinkesteijn J. *Nederlands Gezondheidszorgonderzoek bij somatische chronische ziek(t)en 1980–1989.* Amsterdam, the Netherlands: Universiteit van Amsterdam, Instituut voor Sociale Geneeskunde; 1991.
5. Spector WD, Katz S, Murphy JB, Fulton JP. The hierarchical relationship between activities of daily living and instrumental activities of daily living. *J Chronic Dis.* 1987;40:481–489.
6. Kempen GIJM, Suurmeijer TPBM. The development of a hierarchical polychotomous ADL–IADL scale for noninstitutionalized elders. *Gerontologist.* 1990;30:497–502.
7. Kempen GIJM. *Thuiszorg voor ouderen.* Groningen, the Netherlands: STYX Publications; 1990.
8. Suurmeijer TPBM, Kempen GIJM. Behavioural changes as an outcome of disease: the development of an instrument. *Int J Health Sci.* 1990;1:189–194.
9. Blaxter M. *The Meaning of Disability. A Sociological Study of Impairment.* London, England: Heinemann; 1980.
10. Locker D. *Disability and Disadvantage. The Consequences of Chronic Illness.* London, England: Tavistock Publications; 1983.
11. Kempen GIJM, Doeglas DM, Suurmeijer TPBM. *Het meten van problemen met zelfredzaamheid op verzorgend en huishoudelijk gebied met de Groningen Activiteiten Restrictie Schaal (GARS). Een handleiding* [Groningen Activity Restriction Scale]. Groningen, the Netherlands: University of Groningen, Northern Centre for Health Care Research; 1993.
12. Briançon S, Doeglas D, Guillemin F, et al. European Research on Incapacitating Diseases and Social Support (EURIDISS). *Int J Health Sci.* 1990;1:217–228.
13. Myers AM. The clinical Swiss army knife. Empirical evidence on the validity of IADL functional status measures. *Med Care.* 1992;30(suppl):MS96–MS111.
14. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.
15. Molenaar IW. Mokken scaling revisited. *Kwantitatieve Methoden.* 1982;3:145–164.
16. Debets P, Brouwer E. *MSP: A Program for Mokken Scale Analysis for Polychotomous Items.* Amsterdam, the Netherlands: University of Amsterdam, Technisch Centrum; 1986.
17. Niemöller K, Van Schuur W. Stochastic models for unidimensional scaling: Mokken and Rasch. In: McKay D, Schofield N, Whiteley P, eds. *Data Analysis and the Social Sciences.* London, England: Frances Pinter Publishers; 1983:120–170.
18. Hunt SM, McKenna SP, McEwen J, Williams J, Papp F. The Nottingham Health Profile: subjective health status and medical consultations. *Soc Sci Med.* 1981; 15A:221–229.
19. Kind P, Carr-Hill R. The Nottingham Health Profile: a useful tool for epidemiologists? *Soc Sci Med.* 1987;25:905–910.
20. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: McLeod CM, ed. *Evaluation of Chemotherapeutic Agents.* New York, NY: Columbia University Press; 1948.
21. Goldberg D, Williams P. *A User's Guide to the General Health Questionnaire.* Windsor, England: NFER Nelson Publishing Company Ltd; 1988.
22. Goldstein MS, Siegel JM, Boyer R. Predicting changes in perceived health status. *Am J Public Health.* 1984;74:611–614.
23. Kempen GIJM, Myers AM, Powell LE. Hierarchicality in ADL and IADL: Analytical Assumptions and Applications for Clinicians and Researchers. Submitted for publication.
24. Guralnik JM, Branch LG, Cummings SR, Curb JD. Physical performance measures in aging research. *J Gerontol.* 1989;44: M141–M146.
25. Linn MW, Hunter KI, Linn BS. Self-assessed health, impairment and disability in Anglo, Black and Cuban elderly. *Med Care.* 1980;18:282–288.
26. Bowling A. *Measuring Health: A Review of Quality Measurement Scales.* Philadelphia, Pa: Open University Press; 1991.
27. Spilker B, ed. *Quality of Life Assessments in Clinical Trials.* New York, NY: Raven Press; 1990.

Functional Transitions among the Elderly: Patterns, Predictors, and Related Hospital Use

Vincent Mor, PhD, Victoria Wilcox, PhD, William Rakowski, PhD, and Jeffrey Hiris, BA

ABSTRACT

Objectives. This paper describes 6-year rates and correlates of functional change in the elderly, as well as associated hospital use.

Methods. The Longitudinal Study on Aging (n = 7527) and matched Medicare claims were used to calculate 6-year functional status transition rates and hospital use rates. A hierarchical measure that incorporated activities of daily living, instrumental activities of daily living, and competing risks of institutionalization and death was used to assess functional status. Multinomial logistic regression was used to predict 1990 status.

Results. The functional status of 12% of men and women 70 to 79 years of age who were initially impaired in instrumental activities of daily living improved, and about half of the initially independent people in that age group remained so. Multivariate analyses revealed that age, baseline functioning, self-rated health, and comorbidity predicted 1990 status. Both baseline functioning and functional change were related to hospitalization.

Conclusions. This study supports others that have shown some long-term functional improvement, but more commonly decline, in the elderly. Furthermore, it documents the link between functional decline and increased hospital use. (*Am J Public Health.* 1994;84:1274–1280)

Introduction

The ability to perform routine daily tasks is an important aspect of quality of life. In general, this ability declines with age.[1,2] In 1985, an estimated 5 million community-dwelling elderly people experienced a functional impairment lasting 90 days or longer; an additional 1.3 million were institutionalized.[3] Physically disabled elderly people face an elevated risk of institutionalization and death.[4]

Disability in the elderly is not always permanent; some experience a return of lost functional capacity.[2,5–8] One purpose of this study was to examine change in functional status in Longitudinal Study on Aging participants. Although several previous studies have examined disability transitions using Longitudinal Study on Aging data,[4,6,9–13] they have used shorter time periods, been confined to subpopulations, examined a limited definition of functioning, or ignored the competing risks of death or institutionalization.

A second purpose of this study was to identify predictors of functional change in the elderly, focusing on age, gender, and self-reported health. Some evidence suggests that age decreases the likelihood of regaining lost functional ability.[2] Most, but not all, of the available data indicate a higher prevalence of disability among women than men,[1–3,14] although research concerning gender differences in the incidence of disability is inconclusive.[2,8,10,15] Positive health perceptions have been associated with a lowered risk of institutionalization, mortality, and functional decline,[7,15–18] and these associations are independent of more objective health indicators.

Little is known about the relationship between functional change and hospital use. Disability levels are positively associated with number of physician visits, length of hospital stay, use of home health services, and likelihood of both hospitalization and institutionalization.[12,19–23] However, it is not clear whether use is a function of disability per se or the process of functional decline. Thus, our third purpose was to assess the relationship between functional change and hospital use and costs.

Methods

Sample

This paper is based on (1) the Longitudinal Study on Aging person-level interview files, (2) Medicare claims files, and (3) a match of Medicare hospital claims with the National Death Index. The Longitudinal Study on Aging uses baseline data from the 1984 National Health Interview Survey (NHIS) Supplement on Aging, which involved the collection of health and social data on NHIS participants 55 years of age and older. The Longitudinal Study on Aging includes 7527 Supplement on Aging participants 70 years old or older who were still living in their respective communities in 1984. These participants were reinterviewed in 1986, 1988, and 1990. A complete description of the Longitudinal Study of Aging has been published previously.[24,25]

SUDAAN Version 6.0[26] was used in weighting and adjusting the sample for

The authors are with the Center for Gerontology and Health Care Research and the Department of Community Health, Brown University, Providence, RI.

Requests for reprints should be sent to Vincent Mor, PhD, Center for Gerontology and Health Care Research, Box G-B220, Brown University, Providence, RI 02912.

This paper was accepted March 16, 1994.

the complex sampling design. All analyses, except where otherwise noted, involved the weighted and adjusted sample.

Measures

Health. Comorbid conditions were assessed by counting the number of serious medical problems reported by the respondent from a list of 13 conditions (e.g., stroke, broken hip, cancer). Self-rated health was assessed by the following standard NHIS question: "Would you say your health in general is excellent, very good, good, fair, or poor?" Response categories were recoded as excellent/very good, good, and fair/poor/don't know (there were some 40 "don't know" cases). "Don't know" responses were included in the latter category because previous research[27] has shown that such responses are associated with an elevated mortality risk comparable to that associated with fair self-rated health.

Functional status. Activities of daily living and instrumental activities of daily living items were used in assessing functional status. Item selection was based on preliminary analyses designed to identify the items that best predicted functional decline and mortality.[10,17] The activities of daily living items assessed respondents' ability to toilet, dress, bathe, transfer, eat, and walk. The instrumental activities of daily living items measured the ability of respondents to prepare meals, shop for personal items, and do light housework. Respondents were asked whether they had difficulty performing each activity, how much difficulty they had, and whether they received help from another person. Respondents who were "unable" or "receiving help" were classified as unable and compared with those who reported being able to perform each activity independently.

Correlations between activities of daily living and instrumental activities of daily living items are high enough to suggest that they measure a single domain, with impairment in activities of daily living more severe than impairment in instrumental activities of daily living.[22] Therefore, we constructed a hierarchical scale in which individuals were classified as (1) independent (able in all activities), (2) impaired in instrumental activities of daily living only (unable in one or more such activities), (3) moderately impaired in activities of daily living (unable in one to two such activities, irrespective of ability in instrumental activities of daily living), or (4) severely impaired in activities of daily living (unable in three or more such activities, irrespective of ability in instrumental activities of daily living). As a result of missing 1984 functional status data, 120 cases were dropped from all analyses. Several categories were added to the 1990 status outcome: institutionalized at the time of the 1990 reinterview attempt, dead, and missing. The dependent variable included a "missing" category to control for the known bias that nonresponse is correlated with disability.[28]

TABLE 1—Characteristics of Study Participants at 1984 Baseline

	Functional Status Data Available (n = 7407), %	Use Data Missing (n = 1406), %	Use Data Available (n = 6001), %	P[a]
Gender				<.01
Male	38.7	24.7	42.0	
Female	61.3	75.3	58.0	
Age, y				<.01
70–74	41.6	35.8	43.0	
75–79	30.6	29.3	31.0	
80–84	16.8	18.6	16.4	
85–99	10.9	16.3	9.6	
Race				NS
White	90.6	89.2	90.9	
Other	9.4	10.8	9.1	
Marital status				<.01
Married	48.2	41.0	49.9	
Never married	4.6	2.6	5.1	
Other	47.2	56.4	45.0	
Living arrangement				=.01
With others	63.8	60.2	64.6	
Alone	36.2	39.8	35.4	
Self-rated health				<.01
Excellent/very good	35.7	31.8	36.6	
Good	30.8	30.2	31.0	
Fair/poor/don't know	33.5	38.0	32.5	
Functional status				<.01
Independent	81.5	73.4	83.4	
IADL disabled	6.6	9.4	5.9	
Moderately ADL disabled	7.3	8.8	7.0	
Severely ADL disabled	4.6	8.5	3.6	

Note. IADL = instrumental activities of daily living; ADL = activities of daily living; NS = not significant.
[a]Use data available vs use data missing.

Institutionalization. To ascertain institutionalization at the time of the interview, we used the National Center for Health Statistics code indicating that subjects were not asked about receiving help with activities of daily living or instrumental activities of daily living because they resided in a nursing home. After those with missing 1984 functional status were eliminated, 262 people were residing in a nursing home at the 1990 interview.

Mortality. Mortality was based on National Death Index matching, verified by information obtained from attempts to conduct the next scheduled interview. When the National Death Index indicated that the respondent was deceased but respondent tracking information indicated that he or she was still alive, we assumed that the interview data were correct.

Hospital use. Medicare Part A and MEDPAR summary data were linked by year to Longitudinal Study on Aging participants who provided their health insurance claim number. We used the hospital claims data to determine the probability of hospitalization in each year and the associated Medicare reimbursement (standardized to 1990 dollars using the Medical Care Consumer Price Index). Respondents who were not matched on both Medicare A and B (n = 2045) or had no record of any Medicare A or B (n = 595) claims filed between 1984 and either 1990 or their death were dropped from the use analyses to avoid artificially deflating the numerator.

TABLE 2—Cross Tabulation of 1984 and 1990 Functional Status, Stratified by Age: Men

	1984 Status, %							
	No Disability		IADL Impairment Only		Moderate ADL Impairment		Severe ADL Impairment	
1990 Status	70–79 y (n = 1932)	80+ y (n = 489)	70–79 y (n = 64)	80+ y (n = 50)	70–79 y (n = 114)	80+ y (n = 67)	70–79 y (n = 54)	80+ y (n = 47)
No disability	48.5	22.4	12.4	0	7.0	2.7	6.9	0
IADL impairment only	3.3	3.3	12.1	5.5	5.5	3.5	0	0
Moderate ADL impairment	3.6	3.8	4.5	3.9	9.9	4.2	3.8	2.0
Severe ADL impairment	1.8	2.8	3.1	6.2	10.2	10.1	6.9	2.0
Institutionalized	1.8	3.3	2.9	2.5	4.1	5.7	6.1	0
Dead	27.4	52.7	47.6	72.3	55.5	66.7	69.3	92.3
Missing	13.6	11.8	17.4	9.6	7.8	7.0	7.1	3.6

Note. IADL = instrumental activities of daily living; ADL = activities of daily living.

TABLE 3—Cross Tabulation of 1984 and 1990 Functional Status, Stratified by Age: Women

	1984 Status, %							
	No Disability		IADL Impairment Only		Moderate ADL Impairment		Severe ADL Impairment	
1990 Status	70–79 y (n = 2735)	80+ y (n = 875)	70–79 y (n = 187)	80+ y (n = 190)	70–79 y (n = 167)	80+ y (n = 201)	70–79 y (n = 105)	80+ y (n = 130)
No disability	53.5	24.6	12.2	2.9	10.0	1.8	3.5	0
IADL impairment only	4.3	5.7	9.8	6.6	3.9	3.4	0	0.7
Moderate ADL impairment	5.7	9.3	11.1	9.0	14.1	7.7	6.2	0.7
Severe ADL impairment	2.3	4.6	5.5	6.0	11.9	8.3	18.1	6.0
Institutionalized	2.6	5.4	4.9	9.6	5.6	9.7	8.0	3.9
Dead	14.0	31.3	36.1	56.6	35.8	56.1	52.4	81.5
Missing	17.6	19.0	20.3	14.2	18.8	13.0	11.8	7.2

Note. IADL = instrumental activities of daily living; ADL = activities of daily living.

Analytic Approach

SUDAAN[26] was used to generate 6-year functional transition tables, stratified by age and sex, from the weighted data. The BMDP[29] multinomial logistic regression procedure was used in weighting the data; however, this procedure did not adjust for the complex sampling as a result of software limitations. The seven-category 1990 status variable was the outcome measure, and 1984 functioning, age, sex, self-reported health, number of serious illnesses, and an Age × Sex interaction were entered as predictors. Hospital use and reimbursements per user were calculated for 4 calendar years—1986, 1987, 1988, and 1989—and are presented here in a weighted format. The tables include actual sample size numbers, but the percentages reflect weighting.

Analyses of functional transitions were based on 7407 individuals with 1984 functional status data. Analyses of hospital use were based on a subsample of 6001 individuals for whom it was possible to match the Medicare health insurance claim number with the Longitudinal Study of Aging personal-level interview file and who had some indication of Part A or MEDPAR use for the 6-year period.

Results

Sample Characteristics

Individuals with valid 1984 functional status data were compared with the 120 individuals with missing 1984 functional status data on the sociodemographic variables, living arrangements, and self-rated health. Those with missing 1984 data had poorer self-rated health but did not differ otherwise from those with valid data.

Table 1 presents descriptive data for three groups at baseline: (1) those with valid 1984 functional status data, (2) those excluded because they could not be matched with Medicare claims data, and (3) those with valid Medicare use and 1984 functional status data. Transitions analyses were based on group 1; use analyses were based on group 3 only. The sample was predominantly female, 70 to 79 years of age, White, currently or previously married, and living with others. All groups reported a mean of six serious medical conditions.

The last two columns of Table 1 summarize chi-square tests comparing the baseline characteristics of those with valid and those with missing use data. Use data were more likely to be missing for those

TABLE 4—Results of the Multinomial Regression Predicting 1990 Functional Status

	1990 Functional Status				
	IADL Disabled OR (95% CI)	Moderately ADL Disabled OR (95% CI)	Severely ADL Disabled OR (95% CI)	Institutionalized OR (95% CI)	Dead OR (95% CI)
IADL disabled in 1984	6.6 (4.7, 9.3)	5.3 (3.8, 7.4)	7.6 (5.2, 11.0)	6.7 (4.6, 9.6)	6.6 (5.1, 8.6)
Moderately ADL disabled in 1984	4.1 (2.8, 6.2)	7.8 (5.6, 11.0)	18.0 (13.0, 26.0)	9.8 (6.8, 14.0)	8.6 (6.6, 11.0)
Severely ADL disabled in 1984	0.68 (0.15, 3.0)	5.9 (3.0, 11.0)	40.0 (22.0, 72.0)	17.0 (9.1, 32.0)	30.0 (18.0, 51.0)
80 years of age or older	2.2 (1.6, 3.2)	2.2 (1.6, 3.2)	3.4 (2.3, 4.8)	3.7 (2.5, 5.5)	4.6 (3.9, 5.5)
Female	1.0 (0.84, 1.3)	1.4 (1.1, 1.6)	1.1 (0.87, 1.4)	1.2 (0.96, 1.6)	0.47 (0.42, 0.52)
Good self-rated health	1.6 (1.3, 2.0)	1.5 (1.3, 1.8)	1.2 (0.91, 1.5)	1.8 (1.4, 2.3)	1.5 (1.4, 1.7)
Other self-rated health	2.5 (2.0, 3.1)	2.1 (1.7, 2.5)	1.8 (1.5, 2.3)	2.7 (2.1, 3.5)	3.0 (2.7, 3.3)
No. of illnesses	1.4 (1.2, 1.5)	1.3 (1.2, 1.5)	1.3 (1.2, 1.4)	1.2 (1.1, 1.3)	1.3 (1.2, 1.3)
Age × gender	1.4 (0.92, 2.2)	1.5 (1.0, 2.2)	1.0 (0.68, 1.6)	1.4 (0.90, 2.2)	1.3 (1.0, 1.6)

Note. The reference group consists of those subjects who were independent in both activities of daily living (ADLs) and instrumental activities of daily living (IADLs) in 1990. OR = odds ratio; CI = confidence interval.

who were female, older, not currently married, and living alone and for those who had poorer self-rated health.

Functional Status Transitions

Tables 2 and 3 present transitional probabilities for functional status over a 6-year interval, stratified by age, for men and women, respectively. Over the 6-year period, stability was not the rule, except for the younger respondents who were independent in 1984. Although functional decline was more common than improvement, approximately 4% of men and 2% of women with known 1990 status improved over the 6-year interval. Most strikingly, 3.5% of the women and 7% of the men 70 to 79 years of age returned to complete independence after being in the most severely impaired condition in 1984.

Despite improvement in a minority of the functionally impaired at baseline, most had died by 1990. Among the survivors, disability increased the likelihood of being in an institution in 1990; nonetheless, the majority of elderly respondents impaired in activities of daily living were residing in the community at the follow-up interview. Men were more likely to die than women, and women were more likely to be institutionalized.

Predictors of Functional Transitions

Table 4 shows the results of the multinomial logistic regression predicting 1990 outcomes while adjusting for age, gender, self-rated health, number of medical conditions, baseline functional status, and the interaction between age and gender. Odds ratios and associated 95% confidence intervals are presented for five outcome states; independence in 1990 is the reference outcome. Although results for those with missing 1990 outcome data are not shown, the model adjusts for this competing risk. The odds ratios for each outcome adjust for the competing risks of being in a different outcome state and are expressed relative to being in the independent category. Because 1984 functional status is entered into the equation, the analysis predicts change from 1984 to 1990.

Functional impairment in 1984 was strongly related to functional independence in 1990, to being in a nursing home, and to having died. Being 80 years of age or older was associated with at least a two-fold risk of being in a nonindependent state in 1990 and more than a fourfold increased risk of death. Gender had a small but statistically significant effect on being moderately disabled but had no effect on the other disability states. While women were much less likely to have died, controlling for other factors, we found no significant effect of gender on being in a nursing home. Health status, measured both as the number of serious comorbidities and as self-rated health, was related to all outcomes.

Functional Transitions, Hospital Use, and Cost

Table 5 shows Medicare-reimbursed hospital use and 1990 costs in relation to functional status transitions. The two groups that were impaired in activities of daily living were combined in these analyses. For each group defined by 1984 and 1990 functional status, Table 5 shows use during 1986, 1987, 1988, and 1989. To illustrate, for elderly people who remained independent from 1984 to 1990, an estimated 13% had at least one hospital visit in 1986, at a mean cost per user of $8957. In 1987, 14% of this group were hospitalized, at a mean cost of $9085 per user.

Several points in Table 5 are noteworthy. First, functional decline was associated with higher hospital use rates. For example, 20% of initially independent persons who were dependent in activities of daily living in 1990 were hospitalized in 1986 (vs 13% among those who remained independent); by 1989, 31% had been hospitalized. Second, those who improved had lower rates of hospitalization than those who remained dependent, particularly as 1990 approached. Indeed, the minority who returned to independence had 1989 hospitalization rates similar to initially independent respondents. Third, proximity to death tended to be associated with an increased probability of hospitalization and higher costs per user.

Discussion

Our results support previous research showing considerable changes in functional status in the elderly. The probability of functional improvement was similar to that found in a recent study.[28] Return to independent functioning following even severe disability did occur. However, stability and improvement were rare among those 80 years old and older; age clearly influences the course of disability and the likelihood of death, even for elderly people with similar baseline functioning.

The fact that a higher percentage of women than men experienced functional decline and institutionalization appears to

TABLE 5—Medicare-Reimbursed Hospital Use and Costs in 1986, 1987, 1988, and 1989, by Participant's Functional Status in 1984 and 1990

1990 Functional Status	1986		1987		1988		1989	
	% Using	Mean Cost per User, $ (SE)	% Using	Mean Cost per User, $ (SE)	% Using	Mean Cost per User, $ (SE)	% Using	Mean Cost per User, $ (SE)
Independent in 1984								
Independent	13	8 957 (615)	14	9 085 (675)	15	9 597 (612)	15	9 159 (559)
IADL dependent	27	8 136 (1 220)	22	10 003 (2 145)	16	11 443 (1 819)	26	8 037 (1 164)
ADL dependent	20	10 693 (1 207)	26	12 003 (1 170)	27	9 942 (1 086)	31	14 628 (2 088)
Institutionalized	29	8 413 (1 129)	29	9 021 (1 310)	34	11 040 (1 460)	31	9 347 (1 159)
Dead	39	38 865 (5 030)	43	31 506 (3 632)	48	42 054 (4 189)	56	67 628 (11 728)
Missing	18	11 038 (1 063)	20	14 535 (1 508)	21	12 474 (1 195)	19	16 655 (1 958)
Total	20	20 622 (1 982)	22	17 686 (1 289)	22	19 582 (1 403)	22	23 896 (2 853)
IADL dependent in 1984								
Independent	7	13 240 (5 333)	16	7 144 (1 744)	27	7 515 (1 018)	17	15 094 (6 794)
IADL dependent	40	5 249 (1 093)	18	4 839 (1 052)	19	7 500 (1 745)	24	5 974 (1 258)
ADL dependent	31	11 720 (1 742)	26	9 204 (2 507)	20	7 653 (1 983)	28	5 160 (908)
Institutionalized	41	12 008 (4 544)	20	14 350 (1 690)	49	11 503 (4 388)	37	7 352 (1 125)
Dead	41	41 770 (11 052)	53	28 892 (9 418)	52	29 848 (7 084)	56	131 307 (75 804)
Missing	26	10 957 (1 642)	26	13 556 (3 782)	18	17 565 (3 803)	28	13 574 (2 923)
Total	34	25 264 (5 627)	33	20 465 (5 310)	31	18 631 (3 527)	32	47 162 (24 698)
Moderately ADL dependent in 1984								
Independent	10	9 621 (3 413)	20	8 899 (3 522)	16	7 776 (1 690)	17	2 875 (292)
IADL dependent	27	7 887 (1 867)	32	9 660 (2 487)	18	9 128 (2 989)	18	3 743 (740)
ADL dependent	33	10 276 (1 974)	23	7 964 (1 211)	21	9 330 (2 007)	29	10 663 (2 241)
Institutionalized	31	8 492 (1 492)	38	8 956 (2 349)	34	8 728 (2 369)	25	6 852 (1 474)
Dead	40	41 307 (9 227)	44	24 051 (3 648)	50	41 232 (11 374)	69	67 278 (18 676)
Missing	31	15 654 (5 723)	31	14 821 (5 472)	27	10 243 (3 017)	20	25 132 (7 136)
Total	33	25 807 (4 458)	33	16 664 (1 928)	31	24 768 (5 648)	33	34 138 (7 957)
Severely ADL dependent in 1984								
Independent	17	8 671 ...	33	24 092 (398)	17	6 110 ...	17	26 487 ...
IADL dependent	...	...	...	...	...	...	...	...
ADL dependent	18	7 881 (3 311)	30	13 639 (3 330)	19	12 910 (5 157)	27	11 351 (2 113)
Institutionalized	49	15 736 (3 867)	20	3 776 (754)	14	2 622 (540)	20	6 434 (2 046)
Dead	39	28 594 (8 048)	54	62 716 (32 732)	50	40 225 (13 815)	61	32 558 (9 128)
Missing	19	13 172 (1 848)	19	6 054 (429)	25	12 552 (3 422)	18	7 591 (2 982)
Total	33	23 529 (5 764)	40	47 040 (22 816)	30	28 478 (8 440)	32	20 393 (4 701)

Note. The denominator is the number of individuals alive and in that status in 1984 and 1990. IADL = instrumental activities of daily living; ADL = activities of daily living.

reflect women's greater likelihood of survival. This conclusion is supported by previous reports that women who become disabled live longer than their male counterparts[3,12,30] but are more likely to enter a nursing home.[4] Our analyses highlight the importance of controlling for the competing risks of dying and institutionalization when investigating the possibility of gender differences in functional change. After adjusting for initial disability and competing risks, we found no consistent effects of gender on changes in physical functioning, corroborating recent reports.[2,7]

Our study extends the literature demonstrating the ability of personal health appraisals to predict future health outcomes by showing that self-rated health predicts subsequent functional status, institutionalization, and death, even when these competing risks are included in a single model. Elderly people who rated their health as less than excellent were more likely to be disabled, institutionalized, or dead in 1990, even after controlling for 1984 functional status and the presence of serious disease.

This study assessed the point prevalence of institutional residence in 1990 rather than the cumulative prevalence of nursing home placement during the 6-year interval. Because nursing home episodes that ended before 1990 were not included in our measure, our results underestimate the 6-year incidence of nursing home placement. Many individuals have relatively short nursing home episodes before returning home or dying.[31–33]

To our knowledge, this is the first study to examine hospital use in association with functional status transitions. Elderly respondents who were more disabled at baseline showed higher use rates and costs during each time interval examined. The costs per user increased over time for each group defined by initial functional status. This pattern of increasing costs was most notable among those who died. These findings are consistent with a report from Health Care Financing Administration analysts who examined Medicare payments for cohorts over 16 years.[34]

Individuals whose functional status improved showed a reduced likelihood of hospitalization in future years. This suggests that temporary impairment is associated with a transient elevation in health care costs that presumably dissipates once the medical cause of the impairment has resolved. Among elderly respondents whose status declined, use increased for those who had no impairment at baseline but not for those with preexisting impairment in 1984. Thus, it appears that the emergence of new disability in previously independent elderly people, rather than the exacerbation of impairment in those previously disabled, leads to an increase in use. However, the temporal sequencing of our data dictates that any conclusions about causal relationships be viewed as tentative.

Our study is limited by the fact that we did not consider intermittent transitions during the 6-year period. However, the consistent level of fluctuation observed in both 2- and 4-year periods across multiple data sets suggests that there are probably real transitions in health and functioning as well as more transitory fluctuations that are at least partially due to acute states that affect functioning.[16,28] Future research should examine the periodicity of these fluctuations to improve current transition rate models.

Another limitation is the high and biased rate of failure to match claims data. Those without matched Medicare data were older, in poorer self-rated health, and more likely to be female. These factors are associated with greater use[35]; thus, our analyses probably underestimate use. In contrast, our decision to exclude matched cases without any claims may have biased toward overuse, but analyses comparing claims and self-reported hospitalizations in these data suggest that the claims may be incomplete, thus justifying our decision.[36] Future studies should use more complete measures of use. Furthermore, since functional status is largely the outcome of multiple disease and age-related processes, our analyses cannot address the "true" causes of hospitalization.

In summary, our analyses indicate some long-term functional improvement, even among those in their 70s. However, decline and death were more prevalent, particularly in men 80 years of age and older and among the elderly with a preexisting disability. Illness and perceived health are robust predictors of all future outcome states, even more so than age and nearly as much as function. Finally, as has been long suspected, functional change is strongly related to future hospital use, with decreased use associated with improvement and increased use associated with decline, particularly as death approaches. □

Acknowledgments

Preparation of this paper was supported by Agency for Health Care Policy Research grant HS 06795.

References

1. Jette AM, Branch LG. The Framingham Disability Study: II. Physical disability among the aging. *Am J Public Health.* 1981;71:1211–1216.
2. Rogers RG, Rogers A, Belanger A. Disability-free life among the elderly in the United States. *J Aging Health.* 1992;4:19–42.
3. Manton KG. Epidemiological, demographic, and social correlates of disability among the elderly. *Milbank Q.* 1989; 67(suppl 2, pt 1):13–58.
4. Worobey JL, Angel RJ. Functional capacity and living arrangements of unmarried elderly persons. *J Gerontol Soc Sci.* 1990;45: S95–S101.
5. Branch LG, Katz S, Kniepmann K, Papsidero JA. A prospective study of functional status among community elders. *Am J Public Health.* 1984;74:266–268.
6. Crimmins EM, Saito Y. Getting better and getting worse. Transitions in functional status among older Americans. *J Aging Health.* 1993;5:3–36.
7. Kaplan GA, Strawbridge WJ, Camacho T, Cohen RD. Factors associated with change in physical functioning in the elderly: a six-year prospective study. *J Aging Health.* 1993;5:140–153.
8. Manton KG. A longitudinal study of functional change and mortality in the U.S. *J Gerontol Soc Sci.* 1988;43:S153–S161.
9. Harris T, Kovar MG, Suzman R, Kleinman JC, Feldman JJ. Longitudinal study of physical ability in the oldest-old. *Am J Public Health.* 1989;79:698–702.
10. Mor V, Murphy J, Masterson-Allen S, et al. Risk of functional decline among well elders. *J Clin Epidemiol.* 1989;42:895–904.
11. Verbrugge LM. Disability transitions for older persons with arthritis. *J Aging Health.* 1992;4:212–243.
12. Wolinsky FD, Callahan CM, Fitzgerald JF, Johnson RJ. Changes in functional status and the risks of subsequent nursing home placement and death. *J Gerontol Soc Sci.* 1993;48:S94–S101.
13. Yelin EH, Katz PP. Transitions in health status among community-dwelling elderly people with arthritis. A national, longitudinal study. *Arthritis Rheum.* 1990;33:1205–1215.
14. Foley DJ, Berkman LF, Branch LG, Farmer ME, Wallace RB. Physical functioning. In: Cornoni-Huntley J, Brock DB, Ostfeld AM, Taylor JO, Wallace RB, eds. *Established Populations for Epidemiologic Studies of the Elderly.* Washington, DC: National Institute on Aging; 1986:56–94.
15. Branch LG, Ku L. Transition probabilities to dependency, institutionalization, and death among the elderly over a decade. *J Aging Health.* 1989;1:370–408.
16. Idler EL, Kasl S. Health perceptions and survival: do global evaluations of health status really predict mortality? *J Gerontol Soc Sci.* 1991;46:S55–S65.
17. Rakowski W, Mor V, Hiris J. The association of self-rated health with two-year mortality in a sample of well elderly. *J Aging Health.* 1991;3:527–545.
18. Wilcox V, Kasl SV, Idler E. Self-rated health predicts recovery of physical functioning in older people after hospitalization. Presented at the Annual Scientific Meeting of the Gerontological Society of America, November 1992, Washington, DC.
19. Branch LG, Jette AM. A prospective study of long-term care institutionalization among the aged. *Am J Public Health.* 1982;72:1373–1379.
20. Branch L, Jette A, Evashwick C, Polansky M, Rowe G, Diehr P. Toward understanding elders' health service utilization. *J Community Health.* 1981;7:80–92.
21. Hanley RJ, Wiener JM. Use of paid care by the chronically disabled elderly. *Res Aging.* 1991;13:310–333.
22. Spector WD, Katz S, Murphy JB, Fulton JP. The hierarchical relationship between activities of daily living and instrumental activities of daily living. *J Chronic Dis.* 1987;40:481–489.
23. Wolinsky FD, Johnson RJ. The use of health services by older adults. *J Gerontol Soc Sci.* 1991;46:S345–S347.
24. Fitti JE, Kovar MG. The Supplement on Aging to the 1984 National Health Interview Survey. *Vital Health Stat [10].* 1987;1. DHHS publication PHS 87-1323.
25. Kovar MG, Fitti JE, Chyba MM. The Longitudinal Study of Aging: 1984–1990. *Vital Health Stat [10].* 1992;1. DHHS publication PHS 92-1304.
26. *SUDAAN: Professional Software for Survey Data Analysis.* Research Triangle Park, NC: Research Triangle Institute; 1989.
27. Rakowski W, Mor V. Answering "don't know" (DK) to self-assessment questions about health status and physical activity: Associations with subsequent mortality. Presented at the 120th Annual Meeting of the American Public Health Association, November 1992, Washington, DC.
28. Manton KG, Corder LS, Stallard E. Estimates of change in chronic disability and institutional incidence and prevalence rates in the U.S. elderly population from the 1982, 1984, and 1989 National Long Term Care Survey. *J Gerontol Soc Sci.* 1993;48: S153–S166.
29. *BMDP Statistical Software Manual.* Berkeley, Calif: University of California Press; 1990;2.
30. Katz S, Branch LG, Branson MH, Papsidero JA, Beck JC, Greer DS. Active life expectancy. *N Engl J Med.* 1983;309:1218–1224.
31. Keeler EB, Kane RL, Solomon DH. Short-

and long-term residents of nursing homes. *Med Care.* 1981;19:363–370.
32. Liu K, Manton K. Nursing home length of stay and spenddown in Connecticut, 1977–1986. *Gerontologist.* 1991;31:165–173.
33. Lewis MA, Cretin S, Kane RL. The natural history of nursing home patients. *Gerontologist.* 1985;25:382–388.
34. Gornick M, McMillan A, Lubitz J. A longitudinal perspective on patterns of Medicare payments. *Health Aff.* 1993;12:140–150.
35. Wolinsky FD, Coe RM. Physician and hospital utilization among noninstitutionalized elderly adults: an analysis of the Health Interview Survey. *J Gerontol.* 1984;39:334–341.
36. Stearns SC, Kovar MG, Hayes K, Koch GG. Reconciling respondent reports and Medicare claims for national estimates of hospital use. Presented at the meeting of the American Statistical Association, August 1993, San Francisco, Calif.

New Toll-Free Hotline of Resources on Domestic Violence

With domestic violence at epidemic proportions in the United States, the Family Violence Prevention Fund has established a hotline to provide health professionals with resources to better understand and treat victims of spouse abuse. By calling the Health Resource Center On Domestic Violence at 1-800-313-1310, health professionals can request data and materials designed specifically for their specialty, technical assistance for developing domestic violence programs and protocols, and advice from specialists. Health researchers, journalists, and local, state, and federal policymakers also can request information and assistance through the hotline, which is funded by a grant from the US Department of Health and Human Services.

The hotline operates weekdays from 9 AM to 5 PM Pacific time. Callers can request the following:

- Resource materials, including general information on strengthening health care providers' response to domestic violence, and materials specifically designed for a variety of health care specialties
- Technical assistance, including help developing hospital- or clinic-based domestic violence programs and protocols
- Library services, including advice from information specialists who can provide customized computer literature searches, research studies, and other published materials from the nation's most extensive domestic violence collection

A collaborative effort between the Family Violence Prevention Fund and the Trauma Foundation, this special service library maintains a collection of journal articles, reports, surveys, government documents, curricula, and protocols on domestic violence. Library staff have access to MEDLINE and numerous other computer databases, electronic mail, and the Internet. The library is located at the Pacific Center for Violence Prevention at San Francisco General Hospital. In order to cover costs, the library will charge a photocopying fee and postage for materials provided.

The hotline is one of a series of Family Violence Prevention Fund activities. Later this year, the fund will release the results of a national survey assessing whether nurse managers and physician directors at hospital emergency departments around the country are trained to identify and treat victims of domestic violence. With support from the Advertising Council, the fund will also launch a national public education campaign to promote the prevention of domestic violence.

The Outcomes of Patients Newly Admitted to Nursing Homes after Hip Fracture

ABSTRACT

Objectives. The outcomes of elderly, hospitalized patients discharged to nursing homes after hip fracture were examined.

Methods. For 2624 hip fracture patients admitted to any of 43 proprietary nursing homes between 1984 and 1988, admission assessments were examined in relation to 1-month outcomes.

Results. Mean patient age was 82 ± 7 y; 85% of the sample were female. Within 1 month after discharge, 24% had returned home, 12% had been rehospitalized, 3% had died, and 61% remained in the nursing home. Characteristics significantly associated with mortality included disorientation, functional dependency, neurologic diagnoses, and use of cardiac medications, antidepressants, or narcotics. Rehospitalization was significantly associated with age, gender, living with someone, being ambulatory, and functional dependency. Returning home was associated with younger age, living with someone, being ambulatory, and having no disorientation, functional dependency, or psychiatric or neurologic diagnoses, nor any pressure sores.

Conclusions. Better-functioning persons and those with social support returned home; physically and cognitively impaired persons and those taking narcotics, cardiac medications, or antidepressants were likely to die; and younger men, those with social support, those with functional dependency, and those who were free of disorientation were more likely to be rehospitalized. (*Am J Public Health.* 1994;84:1281–1286)

Douglas P. Kiel, MD, MPH, Ann Eichorn, MS, Orna Intrator, PhD, Rebecca A. Silliman, MD, PhD, and Vincent Mor, PhD

Introduction

Since the advent of prospective payment to hospitals, the recuperative phase of treatment for acute conditions such as hip fracture has shifted to the nursing home. Hospital length of stay for hip fracture patients has decreased by 4 to 9 days,[1–4] the number of in-hospital physical therapy sessions has decreased by as much as 50%,[1–3] the number of patients discharged to nursing homes has more than doubled,[2] and the percentage of patients still in nursing homes 6 months after hospital discharge has increased from 13% to 39%.[2] This shifting of the rehabilitative burden from the acute care hospital to the nursing home makes it particularly important for skilled nursing homes to identify patient characteristics associated with either successful rehabilitation or deterioration.

Two recent but small (n's = 114 and 151) prospective studies have examined patient- and setting-specific characteristics associated with returning home vs staying in the nursing home.[5,6] To expand upon these studies, we examined determinants of all potential outcomes (returning home, returning to hospital, dying, or remaining in the nursing home) in patients newly admitted to skilled nursing homes following hip fracture. We hypothesized that individuals with more comorbidity, those who were more functionally dependent, disoriented, immobile, and taking more drugs, would be less likely to return home and more likely to return to the hospital or to die.

Methods

Study Population

Data used in this study come from regular assessments of residents in 43 continuously owned facilities of the National Health Corporation located in Missouri (n = 9), Kentucky (n = 4), South Carolina (n = 5), and Tennessee (n = 25). The sample comprised all residents 65 years of age and older whose primary insurer was Medicare and who had been newly admitted to one of the nursing homes from a hospital between 1984 and 1988 with a primary *International Classification of Diseases,* Ninth Revision, Clinical Modification[7] (ICD-9-CM) diagnosis of hip fracture (ICD-9-CM codes 820.00, 820.02, 820.03, 820.09, 820.20, 820.21, 820.22, and 820.8). These new admissions had never been residents of a National Health Corporation facility, but 38.3% had been institutionalized elsewhere before their hip fractures. The National Health Corporation maintains a comprehensive computerized clinical database of resident characteristics. Data are collected in monthly assessments of all Medicare-covered residents by trained nurses, submitted to a central office for computerization, and reviewed for completeness by corporate staff. Assessments are used within each facility to monitor resident status and to allocate staffing

Douglas P. Kiel is with the Hebrew Rehabilitation Center for Aged Research and Training Institute and Harvard Medical School, Boston, Mass. Ann Eichorn and Vincent Mor are with the Brown University Center for Gerontology and Health Care Research, Providence, RI. Orna Intrator is with the Department of Statistics, Hebrew University, Jerusalem, Israel. Rebecca A. Silliman is with The Health Institute, New England Medical Center and Tufts University School of Medicine, Boston, Mass.

Requests for reprints should be sent to Douglas P. Kiel, MD, MPH, Hebrew Rehabilitation Center for Aged Research and Training Institute, 1200 Centre St, Boston, MA 02131.

This paper was accepted March 30, 1994.

TABLE 1—Baseline Characteristics of the Sample (n = 2624), by Outcome

	All Patients	Still in Nursing Home	Home	Hospital	Dead
Mean age (SD)	83 (7)	83 (7)	81 (7)	82 (7)	85 (7)
Female, %	83	84	84	75	72
Living arrangements, %					
Institutionalized	38	41	27	42	45
Noninstitutionalized	62	59	73	58	55
Able to walk alone or with assistance, %	29	26	38	31	13
Able to perform all ADLs[a], %	59	56	78	48	24
Clear in orientation, %	58	54	78	46	23
No. of secondary diagnoses, %					
0–2	29	28	35	26	22
3–4	34	35	32	32	42
5–8	37	37	33	42	36
Diagnostic categories[b], %					
Infectious disease	7	7	5	9	11
Psychiatric	19	21	12	22	28
Neurological	25	29	17	26	21
Cardiac	49	49	49	50	50
Pressure ulcer, %	20	21	14	25	33
No. of medications, %					
0–3	30	31	30	26	37
4–6	49	50	47	49	36
≥7	22	19	23	25	26
On cardiac medications, %	34	32	32	41	53
On antidepressants, %	5	4	6	5	10
On narcotics, %	17	16	20	16	23

[a]Activities of daily living (ADLs) include bathing, dressing, toileting, transfer, eating/feeding, and grooming.
[b]Diagnostic categories were created by grouping ICD-9-cm diagnoses listed as secondary diagnoses.

resources. Priority is given to obtaining high-quality data, and a 10% sample is audited quarterly by a second specially trained nurse. The average percentage disagreement across all homes between the two nurse assessors was 9% for the subset of variables examined quarterly for reliability (mobility, activities of daily living, continence, pressure ulcers, selected ICD-9 codes, orientation, and behavior) (phone communication, Julia Powell, RN, Vice President of Patient Services, National Health Corporation).

Many National Health Corporation homes admit hospital patients for rehabilitation following an acute illness such as a hip fracture or stroke. Accordingly, these facilities have a higher percentage of residents covered by Medicare (57.2%) than the national average of 11.5% of hip fracture admissions.[8]

Definition of Variables

Variables for this study were derived from resident assessments performed within 7 days of admission to the homes. Independent variables fell into three broad categories: sociodemographic, functional, and medical severity. On the basis of previous literature, variables were chosen from the complete database that might, a priori, be associated with early (within 1 month) transitions from the nursing home. When two variables captured similar information (e.g. "cardiac diagnoses" and "cardiac medications"), we chose one of them. Sociodemographic variables included age, sex, and living situation prior to fracture (alone vs with another person). Functional variables included ability to walk alone or with help (yes/no), orientation (disoriented/occasionally disoriented/clear), and dependency in activities of daily living (bathing, dressing, toileting, transferring, eating). To account for medical disease severity, we postulated a priori that certain categories of secondary diagnoses might influence short-term outcomes in hip fracture patients newly admitted to nursing homes. ICD-9-CM diagnoses were grouped into four such categories: psychiatric diagnoses, cardiac diagnoses, infectious disease diagnoses, and neurological diagnoses. For example, psychiatric diagnoses present in the study population included "organic psychosis," "schizophrenia," "psychosis," "paranoia," "organic brain damage," and "other mental disorders." We also examined commonly used medications such as cardiac medications, antidepressants, and narcotics. Finally, we categorized patients as to whether or not they had pressure ulcers at admission.

For each subject, outcome status (returned home, remained in nursing home, returned to hospital, died) was examined at 1 month. Since our interest was in identifying characteristics associated with the first transition within 30 days of admission, we did not track subjects past the first transition. Thus a person discharged to a hospital and dying in the hospital within 30 days of nursing home admission would be categorized as a discharge to hospital.

Statistical Analysis

Bivariate associations between independent variables and the four possible outcomes were examined by means of chi-square tests of association. All bivariate associations between independent variables and the four possible outcomes that were significant ($P \leq .10$) were fitted to a polychotomous logistic regression model, a generalization of logistic regression to more than two outcome categories (e.g., home, nursing home, hospital, death). Collapsing any of the four outcome states would obscure important relationships. A polychotomous model assumes that the possible outcomes compete with each other and are mutually exclusive. The parameters of the model were estimated by maximizing a likelihood function by means of the mlogit procedure in STATA Release 3.[9] To interpret coefficients of independent variables in a polychotomous model, one of the outcomes is assigned as a referent category; for this analysis, the referent category was remaining in the nursing home. The adjusted relative risk of an outcome of interest for a given independent variable represents the risk (adjusted for other variables in the model) of experiencing that outcome instead of the referent outcome, given the competing risks of experiencing other possible outcomes. This analysis was repeated after stratifying by living arrangement prior to hip fracture (institutionalized vs noninstitutionalized), since these two groups would be expected to have distinctly different premorbid profiles.

Results

Sample Description

Between 1984 and 1988, 2624 Medicare patients were newly admitted to a National Health Corporation nursing home with a primary diagnosis of hip fracture. The mean age of the sample was 83 (± 7) years, 83% were female, and 38% were living in an institution at the time of their fractures (Table 1). Only 29% were able to walk alone or with assistance, but 59% of the sample were able to perform all activities of daily living. More than half of the subjects were clear in their orientation. The number of secondary diagnoses (out of a possible total of eight) was equally distributed across groupings of 0 through 2, 3 or 4, and 5 through 8. Cardiac diagnoses were listed most often as secondary diagnoses, followed by neurological, psychiatric, and infectious disease diagnoses. Approximately 20% of the sample had pressure ulcers.

Table 2 displays the 1-month transition of the original cohort of 2624 patients and the cohort stratified as to prehospital living situation. Within the first month, 23.7% had returned home, 12.4% had been sent back to an acute care hospital, 3.1% had died, and 60.8% were still in the nursing home. Previously institutionalized hip fracture patients were less likely than previously noninstitutionalized persons to return home (17% vs 28%; $P < .05$).

TABLE 2—First Transition within 30 Days of Admission of Patients Newly Admitted to Nursing Homes with Hip Fractures

Transition	All Patients (n = 2624)	Previously Institutionalized* (n = 1005)	Previously Noninstitutionalized* (n = 1619)
Home, %	23.7	16.9	27.9
Still in nursing home, %	60.8	65.8	57.7
Hospital, %	12.4	13.6	11.6
Dead, %	3.1	3.7	2.8

*$P < .0001$ for the difference in outcomes between previously institutionalized and noninstitutionalized patients.

TABLE 3—Polychotomous Logistic Regression Analysis[a] for Entire Study Population (n = 2556)

	Home		Hospital		Death	
Baseline Characteristic	OR	95% CI	OR	95% CI	OR	95% CI
Age, y						
65–74	1.00	Referent	1.00	Referent	1.00	Referent
74–85	0.75	0.56, 1.00	0.58	0.40, 0.83	0.77	0.33, 1.83
>85	0.50	0.37, 0.68	0.55	0.38, 0.80	1.55	0.67, 3.56
Female sex	0.85	0.64, 1.12	0.63	0.46, 0.86	0.61	0.35, 1.06
Living with someone	1.64	1.33, 2.01	1.44	1.12, 1.87	1.30	0.81, 2.10
Any ADL dependency[b]	0.62	0.48, 0.79	1.45	1.08, 1.93	2.23	1.22, 4.11
Disorientation						
None	1.00	Referent	1.00	Referent	1.00	Referent
Occasional	0.62	0.48, 0.81	1.52	1.13, 2.05	2.51	1.33, 4.73
Disorientation/coma	0.30	0.19, 0.48	1.05	0.70, 1.58	3.77	1.88, 7.59
Able to walk (alone or with help)	1.35	1.09, 1.68	1.54	1.16, 2.05	0.76	0.38, 1.52
Pressure ulcer	0.73	0.56, 0.96	1.09	0.81, 1.46	1.37	0.83, 2.25
No. of medications						
0–3	1.00	Referent	1.00	Referent	1.00	Referent
4–6	0.99	0.78, 1.26	1.13	0.83, 1.54	0.57	0.33, 1.01
≥7	1.23	0.90, 1.66	1.43	0.98, 2.09	0.78	0.40, 1.51
Cardiac medications	0.98	0.78, 1.23	1.39	1.06, 1.82	2.33	1.42, 3.84
Antidepressants	1.31	0.85, 2.03	1.08	0.61, 1.93	3.12	1.32, 7.39
Narcotics	1.11	0.86, 1.44	1.03	0.73, 1.44	1.86	1.05, 3.31
Infectious disease diagnoses	0.92	0.60, 1.40	1.27	0.82, 1.97	1.31	0.62, 2.79
Psychiatric diagnoses	0.75	0.55, 1.02	0.92	0.67, 1.27	0.78	0.45, 1.35
Neurological diagnoses	0.66	0.51, 0.86	0.75	0.56, 1.00	0.41	0.23, 0.73

Note. OR = adjusted odds ratio for 1-month outcome; CI = confidence interval. Sample size is not 2624 because of missing data for some variables in the model.

[a]"Staying in nursing home" was the referent outcome.

[b]Activities of daily living (ADLs) include bathing, dressing, toileting, transferring, eating/feeding, and grooming.

Polychotomous Logistic Regression Results

We included all of the above independent variables in a polychotomous logistic regression model except cardiac diagnoses ($P = .97$, bivariate) and number of secondary diagnoses (collinear with individual diagnostic categories), since all bivariate associations were highly statistically significant (data not shown). As shown in Table 3, after all other variables in the model were controlled, older age reduced the probability of returning home or to hospital. Being female reduced the likelihood of returning to hospital and tended to lower the risk of mortality (not statistically significant at $P < .05$). Living with someone increased the probability of returning home and of being rehospitalized. Having any dependency in activities of daily living strongly increased the risks of death, rehospitalization, and not being able to return home. Disorientation was the strongest predictor of death and greatly decreased the probability of returning home. The ability to walk slightly increased the likelihood of returning home or to hospital but was not significantly associated with death. The presence of a pressure ulcer significantly reduced the probability of returning home, but was not associated with a significantly higher risk of death. When the total number of medications, as well as all other variables in the model, were controlled, the use of cardiac medications, antidepressants, and narcotics contributed independently to a higher risk of dying. Having secondary neurological diagnoses reduced the likelihood of rehospitalization, returning home, and death.

When the analysis was repeated after stratifying by living arrangement prior to hip fracture, the results were different (Tables 4 and 5). For previously noninstitutionalized persons, the model was similar to that of the overall sample, although slightly greater effect sizes were observed for the association between disorientation, cardiac medications, antidepressants, and death, whereas narcotic use

TABLE 4—Polychotomous Logistic Regression Analysis[a] for Previously Institutionalized Persons (n = 989)

Baseline Characteristic	Home		Hospital		Death	
	OR	95% CI	OR	95% CI	OR	95% CI
Age, y						
65–74	1.00	Referent	1.00	Referent	1.00	Referent
74–85	0.73	0.42, 1.26	0.51	0.29, 0.89	0.56	0.16, 1.90
>85	0.44	0.24, 0.79	0.50	0.28, 0.89	1.04	0.32, 3.32
Female sex	0.68	0.41, 1.12	0.61	0.39, 0.97	0.58	0.26, 1.28
Any ADL dependency[b]	0.54	0.35, 0.83	1.40	0.90, 2.19	3.02	1.03, 8.84
Disorientation						
None	1.00	Referent	1.00	Referent	1.00	Referent
Occasional	0.52	0.33, 0.84	1.05	0.65, 1.69	1.62	0.56, 4.73
Disorientation/coma	0.19	0.08, 0.45	0.98	0.55, 1.73	2.53	0.82, 7.78
Able to walk (alone or with help)	0.97	0.64, 1.48	1.14	0.72, 1.80	0.68	0.22, 2.08
Pressure ulcer	0.66	0.36, 1.20	1.01	0.63, 1.62	1.35	0.63, 2.87
No. of medications						
0–3	1.00	Referent	1.00	Referent	1.00	Referent
4–6	0.88	0.56, 1.37	1.33	0.82, 2.14	0.57	0.25, 1.29
≥7	1.33	0.76, 2.32	1.32	0.73, 2.40	0.71	0.27, 1.90
Cardiac medications	0.89	0.58, 1.37	1.32	0.87, 1.99	1.70	0.81, 3.56
Antidepressants	1.26	0.58, 2.75	1.14	0.50, 2.61	1.72	0.36, 8.28
Narcotics	1.75	1.09, 2.81	1.15	0.67, 1.99	1.95	0.79, 4.82
Infectious disease diagnoses	1.29	0.67, 2.49	1.24	0.66, 2.33	1.93	0.73, 5.07
Psychiatric diagnoses	0.51	0.29, 0.91	0.80	0.50, 1.27	0.99	0.46, 2.14
Neurological diagnoses	0.73	0.46, 1.18	0.87	0.57, 1.32	0.68	0.31, 1.48

Note. OR = adjusted odds ratio for 1-month outcome; CI = confidence interval.
[a]"Staying in nursing home" was the referent outcome.
[b]Activities of daily living (ADLs) include bathing, dressing, toileting, transferring, eating/feeding, and grooming.

was no longer significant, probably because of a smaller sample size. For previously institutionalized persons, the association between dependency in activities of daily living and death was stronger. Disorientation did not significantly increase the risk of death; however, it reduced the probability of returning home more for the previously institutionalized than for the previously noninstitutionalized. The ability to walk, alone or with help, did not increase the likelihood of returning to the hospital, as it had for noninstitutionalized persons. The use of cardiac medications or antidepressants was not associated with mortality in the previously institutionalized group; however, the use of narcotics did increase the probability of returning home for persons in this group.

Discussion

One month after admission to a nursing home following hip fracture, 24% of elderly Medicare patients had returned home, 61% were still in the nursing home, 12% had been sent back to the hospital, and 3% had died. The mortality rate in this group was slightly less than the rates of 6% to 12% for all hip fracture patients reported by others,[10-13] but closer to the 4% annual mortality for nursing home residents reported by Keene.[10] Entry characteristics independently associated with the first transition within 30 days included age, sex, premorbid living arrangement, functional status, orientation, mobility, neurological diagnoses, and the use of cardiac medications, antidepressants, and narcotics. Results for the previously institutionalized group differed in some instances from those for the previously noninstitutionalized. The ability to walk did not increase the likelihood of returning home for the previously institutionalized, as it had for the previously noninstitutionalized, whereas narcotic analgesic use was associated with returning home only in the previously institutionalized. The hypothesis that risk factors associated with different outcomes would overlap was only partially supported by our findings. With the exception of disorientation and dependency in activities of daily living, admission characteristics differed in their association with the four outcomes.

An important feature of our study is the focus on simultaneously considered outcomes in a large sample of residents from multiple nursing homes. Previous studies of hip fracture patients treated in nursing homes involved relatively small numbers of residents in a limited number of facilities, and the studies focused only on patients' returning home. Our data clearly reveal that this simplification is gained at the expense of precision. Had we focused only on patients' returning home, the contrast between remaining in the nursing home and dying or being rehospitalized would have been lost. Furthermore, combining these outcomes would yield an incorrect estimate of the effect of each independent variable. Consequently, all estimates of the effect of functioning or medical morbidity on given outcomes must be conditioned upon consideration of all the mutually exclusive outcomes.

Previous studies have demonstrated a strong relationship between dependency in activities of daily living and mortality among community-dwelling elders[15] and among persons discharged from nursing homes,[16,17] but to our knowledge this is the first study to describe, by means of multivariate models, the association between functional status and mortality during the first 30 days of nursing home residence. Our finding that functional dependency decreases the probability of returning home has also been reported by others.[5,10] Although we considered functional status at time of admission to the nursing home, rather than prior to fracture, our results were consistent with those of Keene and Anderson.[10] The finding that functional status and the ability to walk had a modest positive effect on rehospitalization as well as on the probability of returning home may imply that these factors not only reflect successful rehabilitation, but also influence the decision to rehospitalize a patient who has an acute decline during rehabilitation.

Disorientation at admission was one of the strongest predictors of mortality, even after other baseline characteristics were controlled. Delirium probably accounts for a significant percentage of residents in this category, since previous work has shown that up to 37% of patients developing delirium in the hospital still have this diagnosis at discharge.[18] The observation that delirium is an important predictor of mortality[12,19,20] may partly reflect the fact that delirium is associated with the presence of significant underlying medical problems. Indeed, an independent effect of delirium on mortality could not be demonstrated in two recent studies that also controlled for illness severity.[18,21] Our finding of an independent effect of

disorientation after controlling for comorbidities and functional status may reflect the medical vulnerability of this condition.

In this study, several diagnoses and medications were associated with 1-month outcomes, possibly serving as comorbidity measures or disease severity measures. For example, neurological secondary diagnoses decreased the probability of all outcomes, and cardiac medications increased both mortality and the likelihood of returning to the hospital. Similarly, the finding that pressure ulcers were associated with a decreased probability of returning home and a trend toward an increase in mortality might be interpreted as simply a reflection of pressure ulcers as markers of underlying medical problems and poor mobility[22]; however, these associations remained after we controlled for functional status, ability to walk, and secondary diagnoses, suggesting an independent effect. Finally, the finding that antidepressant use increased the risk of death more than three times may have arisen because of adverse events associated with the use of antidepressants, such as falls, hip fracture, and cardiac conduction abnormalities,[23] or because the prescribing of these drugs may be selective for those individuals with underlying conditions that are themselves associated with greater mortality. This link would also be a plausible explanation for the strong association between cardiac medication use, narcotic drug use, and mortality.

The unexpected finding that women were less likely than men to be rehospitalized is the first demonstration of an association between sex and hospitalization rates among nursing home residents. The finding that persons with hip fracture who were living with someone prior to fracture were more likely to return home and more likely to be rehospitalized probably reflects the importance of caregiver support at home as well as caregiver advocacy when clinical deterioration occurs.

Our results must be interpreted with consideration of several features. First, this was not a random sample of patients admitted to nursing homes with hip fracture. Second, some of the assessments may be imprecise; however, since the data are collected uniformly and used routinely by the National Health Corporation to assess staffing needs and to improve the efficiency of patient care, these assessments are carefully monitored for accuracy. Since the variables are carefully defined in a coding manual, and since the reliability was high, the chance of misclassification was small. Reviewers were always charge-level nurses on the units who were not motivated by corporate incentives such as reimbursement.

TABLE 5—Polychotomous Logistic Regression Analysis[a] for Previously Noninstitutionalized Persons (n = 1573)

Baseline Characteristic	Home		Hospital		Death	
	OR	95% CI	OR	95% CI	OR	95% CI
Age, y						
65–74	1.00	Referent	1.00	Referent	1.00	Referent
74–85	0.75	0.53, 1.06	0.58	0.36, 0.94	0.97	0.28, 3.38
>85	0.52	0.36, 0.74	0.54	0.33, 0.88	2.19	0.65, 7.43
Female sex	0.85	0.61, 1.19	0.60	0.40, 0.90	0.56	0.26, 1.19
Living with someone	1.35	1.06, 1.73	1.26	0.88, 1.79	1.15	0.57, 2.30
Any ADL dependency	0.66	0.49, 0.90	1.42	0.96, 2.09	2.02	0.94, 4.37
Disorientation						
None	1.00	Referent	1.00	Referent	1.00	Referent
Occasional	0.74	0.54, 1.02	2.01	1.36, 2.97	3.49	1.58, 7.72
Disorientation/coma	0.44	0.25, 0.73	1.11	0.61, 2.05	4.72	1.84, 12.12
Able to walk (alone or with help)	1.59	1.23, 2.05	1.96	1.35, 2.84	0.88	0.36, 2.16
Pressure ulcer	0.77	0.56, 1.05	1.19	0.82, 1.73	1.40	0.71, 2.75
No. of medications						
0–3	1.00	Referent	1.00	Referent	1.00	Referent
4–6	1.06	0.80, 1.41	1.04	0.69, 1.56	0.59	0.27, 1.30
≥7	1.22	0.84, 1.77	1.56	0.94, 2.58	0.80	0.32, 2.03
Cardiac medications	1.06	0.81, 1.39	1.52	1.05, 2.19	3.19	1.59, 6.42
Antidepressants	1.35	0.79, 2.32	1.02	0.46, 2.31	4.40	1.47, 13.13
Narcotics	0.90	0.66, 1.23	0.94	0.61, 1.45	1.77	0.81, 3.90
Infectious disease diagnoses	0.76	0.43, 1.33	1.29	0.70, 2.39	0.98	0.28, 3.49
Psychiatric diagnoses	0.96	0.66, 1.40	1.07	0.68, 1.67	0.71	0.31, 1.64
Neurological diagnoses	0.64	0.47, 0.87	0.66	0.44, 0.98	0.25	0.10, 0.62

Note. OR = adjusted odds ratio for 1-month outcome; CI = confidence interval.
[a]"Staying in nursing home" was the referent outcome.
[b]Activities of daily living (ADLs) include bathing, dressing, toileting, transferring, eating/feeding, and grooming.

Despite these limitations, the findings from this study extend the work of previous studies and capitalize on the large database of the National Health Corporation. Furthermore, our findings are more precise because we considered all outcomes simultaneously. Simplifying the analysis by combining outcomes could lead to a misinterpretation of effects. This method of modeling outcomes would be directly applicable to other conditions requiring rehabilitation in the nursing home, such as stroke. Future studies of such conditions may be able to distinguish patient characteristics that are either generic to rehabilitation in the elderly or disease-specific. □

References

1. Fitzgerald JF, Moore PS, Dittus RS. The care of elderly patients with hip fracture: changes since implementation of the prospective payment system. *N Engl J Med.* 1988;319:1392–1397.
2. Fitzgerald JF, Fagan LF, Tierney WM, Dittus RS. Changing patterns of hip fracture care before and after implementation of the prospective payment system. *JAMA.* 1987;258:218–221.
3. Palmer RM, Saywell RM, Zollinger TW, et al. The impact of the prospective payment system on the treatment of hip fractures in the elderly. *Arch Intern Med.* 1989;149:2237–2241.
4. Ray WA, Griffin MR, Baugh DK. Mortality following hip fracture before and after implementation of the prospective payment system. *Arch Intern Med.* 1990;150:2109–2114.
5. Bonar SK, Tinetti ME, Speechley M, Cooney LM. Factors associated with short- versus long-term skilled nursing facility placement among community-living hip fracture patients. *J Am Geriatr Soc.* 1990;38:1139–1144.
6. Fitzgerald JF, Dittus RS. Institutionalized patients with hip fractures: characteristics associated with returning to community

dwelling. *J Gen Intern Med.* 1990;5:298–303.
7. *International Classification of Diseases,* Ninth Revision, Clinical Modification. Hyattsville, Md: National Center for Health Statistics; 1980. DHHS publication PHS 80-1260.
8. National Nursing Home Survey. *Vital Health Stat [13].* January 1989; no. 97:62–63. DHHS publication PHS 89-1758.
9. *STATA Release 3.* 5th ed. Santa Monica, Calif: Computing Resource Center; 1992.
10. Keene JS, Anderson CA. Hip fractures in the elderly: discharge predictions with a functional rating scale. *JAMA.* 1982;248: 564–567.
11. Ray WA, Griffin MR, Baugh DK. Mortality following hip fracture before and after implementation of the prospective payment system. *Arch Intern Med.* 1990;150: 2109–2114.
12. Magaziner J, Simonsick EM, Kashner M, Hebel JR, Kenzora JE. Survival experience of aged hip fracture patients. *Am J Public Health.* 1989;79:274–278.
13. Weiss NS, Liff JM, Ure CL, Ballard JH, Abbott GH, Daling JR. Mortality in women following hip fracture. *J Chronic Dis.* 1983;36:879–882.
14. Fisher ES, Baron JA, Malenka DJ, et al. Hip fracture incidence and mortality in New England. *Epidemiology.* 1991;2:116–122.
15. Spector WD, Katz S, Murphy JB, Fulton JP. The hierarchical relationship between activities of daily living and instrumental activities of daily living. *J Chronic Dis.* 1987;40:481–489.
16. Lewis MA, Cretin S, Kane RL. The natural history of nursing home patients. *Gerontologist.* 1985;25:382–388.
17. Lewis MA, Kane RL, Cretin S, Clark V. The immediate and subsequent outcomes of nursing home care. *Am J Public Health.* 1985;75:758–762.
18. Levkoff SE, Evans DA, Liptzin B, et al. Delirium: the occurrence and persistence of symptoms among elderly hospitalized patients. *Arch Intern Med.* 1992;152:334–340.
19. Francis J, Kapoor WN. Delirium in hospitalized elderly. *J Gen Intern Med.* 1990;5:65–79.
20. Francis J, Kapoor WN. Prognosis after hospital discharge of older medical patients with delirium. *J Am Geriatr Soc.* 1992;40:601–606.
21. Fields SD, MacKenzie CR, Charlton ME, Sax FL. Cognitive impairment: can it predict the course of hospitalized patients? *J Am Geriatr Soc.* 1986;34:579–585.
22. Bergstrom N, Braden B. A prospective study of pressure sore risk among institutionalized elderly. *J Am Geriatr Soc.* 1992; 40:747–758.
23. Ray WA, Griffin MR, Schaffner W, Baugh DK, Melton LJ III. Psychotropic drug use and the risk of hip fracture. *N Engl J Med.* 1987;316:363–369.

Treatment and Survival among Elderly Americans with Hip Fractures: A Population-Based Study

ABSTRACT

Objectives. This study was undertaken to examine the patterns of treatment and survival among elderly Americans with hip fracture.

Methods. A 5% national sample of Medicare claims was used to identify patients who sustained hip fractures between 1986 and 1989. In comparing treatment patterns across regions, direct standardization was used to derive age- and race-adjusted percentages. Logistic regression and Cox regression were used to examine short- and long-term survival.

Results. In the United States, 64% of femoral neck fractures were treated with arthroplasty; 90% of pertrochanteric fractures were treated with internal fixation. Higher short- and long-term mortality was associated with being male, being older, residing in a nursing home prior to fracture, having a higher comorbidity score, and having a pertrochanteric fracture. Blacks and Whites had similar 90-day postfracture mortality, but Blacks had a higher mortality later on. For femoral neck fracture, internal fixation has a modestly lower short-term mortality associated with it than arthroplasty has.

Conclusion. Variation in the treatment of hip fracture was modest. The increased delayed mortality after hip fracture among Blacks requires further study. (*Am J Public Health.* 1994;84:1287–1291)

Grace L. Lu-Yao, PhD, MPH, John A. Baron, MD, Jane A. Barrett, MSc, and Elliott S. Fisher, MD

Introduction

More than 200 000 hip fractures occur annually in the United States,[1,2] with appreciable associated morbidity and mortality. Despite the human and economic burden of these fractures, relatively little is known about the patterns of treatment or of survival following different treatments.

Although several studies have identified factors related to postfracture survival,[3–6] their results may not be generalizable for several reasons. For example, patient characteristics and surgical practice may vary from one institution to another. The limitations of sample size, the difficulties in comparing crude mortality rates, and the differences in the criteria used to determine treatments further hamper interpretation. To overcome these limitations, we used a 5% sample of national Medicare claims to examine the patterns of treatment for hip fracture and of survival following the most commonly used procedures.

Methods

Identification of Cases

The 5% national Medicare claims database was searched for evidence of hip fracture during the years 1986 to 1989. Both hospital (Part A) and physician (Part B) claims were used. A patient was identified as having an acute hip fracture if he or she had a discharge diagnosis of hip fracture (*International Classification of Diseases,* 9th edition [ICD-9] codes 820.0 to 820.9) or if there was a physician claim indicating a treatment for hip fracture (current procedure terminology [CPT] codes in the range 27230 to 27248). A detailed algorithm for case identification has been described elsewhere[7] and can be obtained from the authors. Patients were excluded from analysis if there was evidence of bone or metastatic cancer, fractures of the femoral shaft or pelvis, or surgery for previous hip fracture. In addition, patients were excluded if they were under age 65, not resident in the United States, or likely to have incomplete claims data because they were not enrolled in both Medicare Parts A and B, were enrolled in a health maintenance organization (HMO), or were receiving coverage through railroad board entitlement.

Definitions of Variables

The site of fracture and treatment were determined from the ICD-9 diagnosis and procedure codes on the hospital claims and from the CPT procedure codes on the physician claims. Fractures were designated as occurring either at the femoral neck or at another site ("pertrochanteric fractures"). Patients receiving internal fixation could be uniquely identified, but since the claims coding could not reliably distinguish between unipolar hemiarthroplasty, bipolar hemiarthroplasty, and total hip replacement, these procedures were grouped together as "arthroplasty." Patients for whom neither

The authors are with the Department of Community and Family Medicine at Dartmouth Medical School, Hanover, NH. In addition, John A. Baron is with the Department of Medicine at Dartmouth Medical School, and Elliott S. Fisher is with the Veterans Administration Medical Center, White River Junction, Vt.

Requests for reprints should be sent to Grace L. Lu-Yao, PhD, MPH, HB 7251, Dartmouth Medical School, Hanover, NH 03755-3863.

This paper was accepted November 3, 1993.

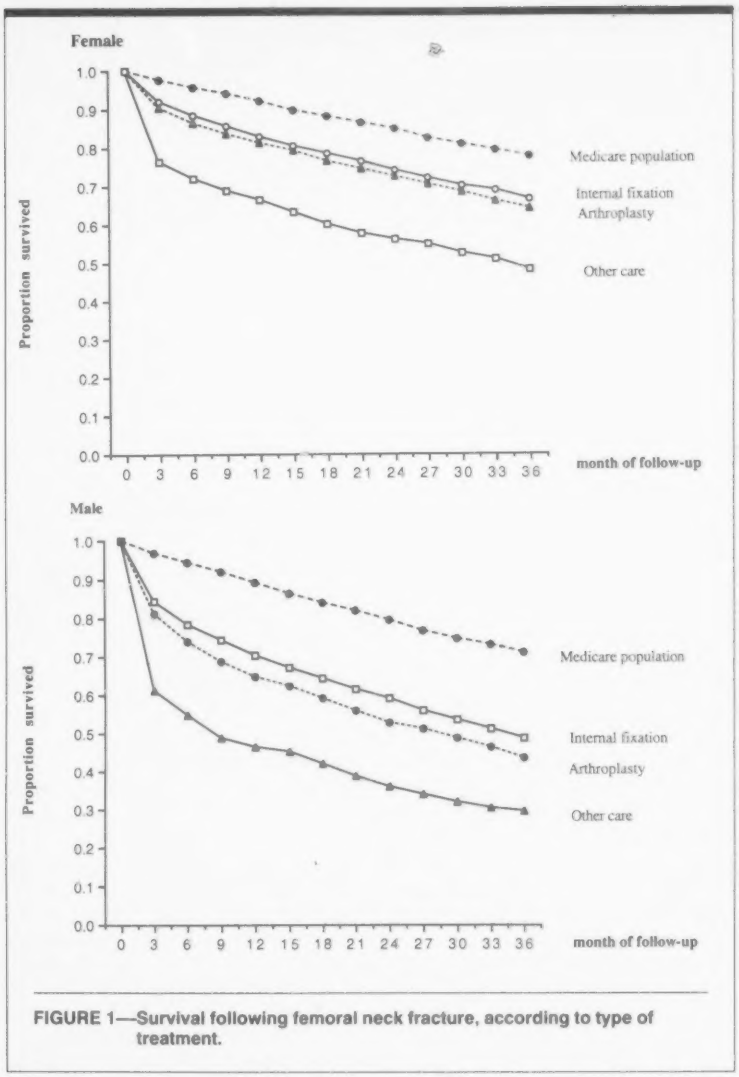


FIGURE 1—Survival following femoral neck fracture, according to type of treatment.

internal fixation nor arthroplasty were coded were considered together as a third treatment category ("other care"). Both hospital and physician claims were used to clarify ambiguity or compensate for missing data from any one source. However, because the accuracy of the hospital claims coding has been investigated and found to be excellent for hip fracture,[8] the hospital designation was used in those cases in which the hospital and physician claims disagreed.

Medicare enrollment status, date of birth, sex, race, residence, vital status as of December 31, 1990, and date of death (if any) were obtained from the Health Care Financing Administration denominator files for 1986 to 1990. A comorbidity score based on the presence or absence of a range of diagnoses recorded in the hospital claims filed for the fracture hospitalization or for hospitalizations during the 6 months prior to hip fracture was calculated using our adaptation[7] of the method described by Charlson et al.[9] A patient was identified as being a nursing home resident if his or her Part B physician claims indicated at least one physician visit occurring in a nursing home in the 3 months preceding the hip fracture. (Federal regulations for skilled and intermediate care facilities require at least this frequency of physician visits for certification.) Since only those who fractured a hip after July 1, 1986, had adequate information on comorbidity and prefracture residence, analyses were restricted to patients with hip fractures occurring between July 1, 1986, and June 30, 1989.

Statistical Analysis

In the investigation of geographic variations in treatment, age- and sex-adjusted proportions of patients receiving a given treatment were calculated for each census division and fracture type, using direct standardization with all patients of the fracture type as the standard population.[10] Logistic regression was used to examine the association between patient characteristics and treatment choice.[11] The reference group (Medicare population) shown in Figure 1 provides a representative survival experience of all those in the 5% Medicare sample at the end of 1987 (midpoint in our time period), with the same sex/race/age structure and inclusion criteria as those for the femoral neck fracture patients. The duration of survival postfracture was calculated from the date of the admission for the hip fracture to the date of death or to December 31, 1990. To distinguish early deaths that might be associated with the treatment, separate analyses were conducted for mortality within 90 days and for survival experience beyond 90 days, using logistic regression[11] and Cox proportional hazards regression,[12] respectively. For the latter, log-log plots of all variables were used to assess the proportional hazards assumptions. To examine whether the effect of treatment on survival varied with time in the Cox regressions, separate proportional hazards models were fitted for several follow-up periods, and the statistical significance of the interaction term between treatment and the logarithm of follow-up time was tested.[12] Because no evidence of time-varying effects beyond 3 months post–hip fracture was found, only the overall relative risks for the late survival analyses are presented.

Covariates included in the models were age, sex, race, prefracture residence, comorbidity, site of fracture, and treatment. To account for the effect of comorbidity, two analytical approaches were used: one included exact comorbidity score in the model, and the other used dichotomized comorbidity status (low comorbidity [Charlson score = 0] vs high comorbidity [Charlson score > 0]). Because both approaches gave very similar results, only the model with dichotomized comorbidity status is presented.

TABLE 1—Characteristics of Patients, by Site of Fracture and Treatment

	Femoral Neck Fracture (n = 13 167)			Other Hip Fracture (n = 13 267)		
	Internal Fixation (n = 3569)	Arthroplasty (n = 8458)	Other Care (n = 1140)	Internal Fixation (n = 11 935)	Arthroplasty (n = 417)	Other Care (n = 915)
Sex[a]: female, %	77.9	80.1	73.7	78.6	76.3	74.5
Race[a]						
White, %	95.1	94.6	91.8	95.3	94.2	94.1
Black, %	3.6	3.6	5.4	2.9	3.8	4.5
Age[b]: mean, y	80.2 ± 0.1	81.1 ± 0.1	81.6 ± 0.2	82.2 ± 0.1	81.8 ± 0.4	82.1 ± 0.3
Prefracture residence[a]: nursing home, %	17.3	16.1	26.0	19.6	17.0	22.1
Comorbidity[b]: mean score	0.68 ± 0.02	0.65 ± 0.01	0.77 ± 0.03	0.67 ± 0.01	0.61 ± 0.05	0.83 ± 0.04

[a]The entries in the table indicate the percentage of patients with certain characteristics among those who have the same site of fracture and treatment.
[b]The entries in the table are the mean and the standard error of the mean.

Results

Case Population

We identified 31 467 candidate hip fracture cases that occurred between July 1, 1986, and June 30, 1989. From those, we excluded 2260 patients who were under age 65, or not resident in the United States, who had evidence of cancer or previous hip surgery, or who had insufficient evidence of hip fracture. Another 2433 patients were excluded because they were not enrollees of both Medicare Parts A and B, were members of HMOs, or were railroad board enrollees. Another 340 cases had multiple hip fracture sites. The remaining 13 167 femoral neck fractures and 13 267 pertrochanteric fractures constituted the basis for this report.

The majority of patients with hip fractures were White women over age 80 (Table 1). Femoral neck fractures and pertrochanteric fractures occurred at similar frequency, but the average age at pertrochanteric fracture was slightly higher than that at femoral neck fracture. The great majority of hip fractures were treated with internal fixation or arthroplasty. Patients treated with other care were more likely to be nursing home residents and to have a higher comorbidity score (Table 1).

Factors Associated with Surgical Treatments

More than half of the femoral neck fractures were treated with arthroplasty (8458/13 167), whereas most pertrochanteric fractures were treated with internal fixation (11 935/13 267) (Table 1). Geographic variation in treatment patterns for both types of fracture was modest. The age- and sex-adjusted proportion of femoral neck fractures treated with arthroplasty ranged from 63% to 68% among the nine census divisions; the adjusted proportion of pertrochanteric fractures treated with internal fixation ranged from 88% to 92%.

Since the great majority of pertrochanteric fractures were treated with internal fixation, the association between patient characteristics and the type of treatment was examined only for femoral neck fractures. Among patients over 65, arthroplasty was more likely to be performed in women, older patients, and those persons not living in nursing homes before hip fracture (Table 2).

TABLE 2—Patient Characteristics and the Relative Likelihood of Being Treated by Arthroplasty among Those with Femoral Neck Fractures (n = 13 167): A 5% National Sample of Medicare Claims, 1986 through 1989

	Odds Ratio[a]	95% Confidence Interval
Female	1.00	. . .
Male	0.84	0.77, 0.91
Low comorbidity	1.00	. . .
High comorbidity	0.95	0.88, 1.02
Non–nursing home	1.00	. . .
Nursing home	0.79	0.72, 0.87
White	1.00	. . .
Black	0.90	0.74, 1.08
Other	1.07	0.81, 1.41
Age, y		
65–69	1.00	. . .
70–74	1.09	0.96, 1.24
75–79	1.20	1.06, 1.35
80–84	1.26	1.12, 1.42
85–89	1.23	1.10, 1.39
90+	1.17	1.03, 1.33

[a]Adjusted for age, sex, race, comorbidity, and place of residence before fracture.

Factors Associated with Mortality

Excess mortality associated with hip fractures was most pronounced in the first several months. The 30-day mortality for men and women was 11% and 6%, respectively, whereas the mortality of an age-matched Medicare population was less than 1%. Taking all patients together, the overall post–hip fracture mortality among Medicare beneficiaries was 7% at 1 month, 13% at 3 months, and 24% at 12 months.

Both short- and long-term mortality were higher among patients who were older, were male, resided in a nursing home, had significant comorbid conditions, or had pertrochanteric fracture (Table 3). After taking these factors into account, there was no significant difference in mortality between Blacks and Whites during the first 90 days postfracture. However, the subsequent mortality was higher in Blacks (rate ratio [RR] = 1.21; 95% confidence interval [CI] = 1.08, 1.36) (Table 3). This pattern was found in all census divisions except New England, which had only 12 Blacks with hip fracture during the study period.

TABLE 3—The Effects of Various Risk Factors on Mortality after Hip Fracture

Risk Factors	0–90 Days[a] Odds Ratio	0–90 Days[a] 95% Confidence Interval	91 Days–3 Years[b] Rate Ratio	91 Days–3 Years[b] 95% Confidence Interval
Site of fracture: pertrochanteric/neck fracture	1.18	1.06, 1.30	1.08	1.02, 1.15
Age: 1-year increase	1.07	1.06, 1.07	1.05	1.04, 1.05
Sex: male/female	2.21	2.04, 2.40	1.65	1.57, 1.74
Race				
Black/White	1.01	0.83, 1.23	1.21	1.08, 1.36
Others/White	0.77	0.57, 1.03	1.65	1.41, 1.94
Comorbidity[c]: high/low	1.89	1.75, 2.04	1.74	1.67, 1.82
Residence: nursing home/non–nursing home	1.39	1.28, 1.52	1.54	1.46, 1.62
Treatment (femoral neck fractures only)				
Arthroplasty/internal fixation	1.21	1.06, 1.38	1.05	0.98, 1.13
Other care/internal fixation	3.24	2.71, 3.87	1.31	1.16, 1.48

Note. For both logistic and proportional hazard models, covariates in the model were age (interval width-1 year), sex, race, comorbidity (low and high), prefracture residence, site of fracture, and treatment.
[a]Calculated using logistic regression.
[b]Calculated using Cox proportional hazard regression.
[c]Comorbidity score greater than or equal to 1 was classified as high comorbidity.

Patients with other, unknown, or missing race also had increased long-term mortality compared with Whites.

For patients with femoral neck fractures, patterns of treatment-specific mortality were examined (Table 3, Figure 1). Patients treated with other care suffered higher mortality than those treated with internal fixation or hemiarthroplasty (Figure 1). Patients treated with arthroplasty had a modestly elevated short-term mortality (RR = 1.21, 95% CI = 1.06, 1.38). However, among those who survived 90 days, there was no detectable difference in mortality beyond 3 months post–hip fracture (Table 3).

Discussion

Based on a national sample of Medicare beneficiaries with hip fracture, we found only modest variation in the treatment across regions and races within the United States. Although there was a difference in post–hip fracture mortality between arthroplasty and internal fixation, the magnitude was small and limited to the first several postfracture months. White patients had the lowest postfracture mortality of the racial groups studied.

Although there appears to be consensus among US surgeons regarding management of femoral neck fractures, the observed treatment pattern differed dramatically from that in some regions of Sweden, where almost all femoral neck fractures receive internal fixation.[13] Because about 75% of femoral neck fractures are displaced,[14] this discrepancy almost certainly reflects differing opinions regarding optimal management for displaced femoral neck fractures.[15,16] These findings underscore this controversy and lend support for an evaluation of the outcomes of these alternative treatments.

The survival patterns we report here are comparable to those in several recently published series: the overall 30-day mortality ranged from 6% to 7% and the 1-year mortality ranged from 22% to 24%.[7,17–19] Several earlier studies have shown better short-term survival in patients with femoral neck fracture than in those with pertrochanteric fractures[5,6,19,20]; however, none of them reached statistical significance, possibly owing in part to limited sample size. Our study confirms that pertrochanteric fracture is associated with an elevated mortality even after adjusting for the effect of age and other risk factors.

It has been reported previously that Blacks have a higher mortality after hip fracture than Whites.[21] We show that these findings are independent of type of fracture and treatment, but only for the long-term mortality following the first 3 months postfracture. Could a difference in the severity of the hip fracture contribute to this difference in survival? Although we did not have information on the degree of trauma, the lack of a mortality differential in the first 90 days suggests that this is unlikely. A more plausible explanation for these long-term differences lies with the factors associated in general with a lower life expectancy among Blacks, such as lower socioeconomic status and limited access to health care.[22]

Our study shows nursing home residents to be less likely to have arthroplasty and more likely to suffer postfracture mortality, associations that were present in all census divisions. It is likely that these findings are owing to the poor health and psychosocial support of nursing home residents. Indeed, when comorbidity is deleted from our regression models, the effect of the nursing home appears substantially stronger. Previous studies have also shown that prefracture health and functional status, as well as psychosocial factors, are associated with outcomes following hip fracture.[23,24]

For femoral neck fractures, patients treated with other care had significantly higher mortality than those treated with arthroplasty or internal fixation. Since almost all patients with hip fractures are treated surgically except those who are severely ill, the elevated mortality among those treated with other care is almost certainly a consequence of patient selection. Survival following internal fixation was slightly better than that following arthroplasty, especially in the first several months postfracture—a finding that is consistent with those of several previous studies.[25–27] These differences could be owing to selection of healthier patients for internal fixation. An alternative explanation is that the higher short-term mortality following arthroplasty may be a consequence of factors such as the longer duration of anesthesia and the greater risk of blood loss during surgery.[26]

The validity of our study depends on the accuracy of the Medicare claims. The possibility of incomplete case ascertainment must be acknowledged, but it is unlikely to have affected our findings materially because hip fracture is among the most reliably coded hospital diagnoses.[8] To further improve the ascertainment, accuracy, and clinical relevance of the data extracted, we used both physician and hospital claims. Moreover, our follow-up was relatively complete. The vital

status of patients was based on Social Security Administration files, which have been found to identify more than 97% of known deaths with a false-positive rate of less than 0.1% among those over age 60.[28]

In summary, this study demonstrates that there is little geographic variation in the treatment for hip fractures within the United States. Differences in mortality between arthroplasty and internal fixation are modest and limited to the short term. Racial differences in survival postfracture seem to reflect general mortality trends. □

Acknowledgments

Support for this work was provided by the National Institute on Aging (AG07146) and the Agency for Health Care Policy and Research (5 R18HS 05745–04).

Part of this work was presented at the national meeting of the American Federation of Clinical Research in Baltimore, Md, May 1992.

The authors are indebted to Dr John E. Wennberg and Dr Robert B. Keller for their valuable advice, and to Dr Thomas Bubolz and Ms Sara Ng for their technical support.

References

1. Lewinnek GE, Kelsey J, White AA, Kreiger NJ. The significance and comparative analysis of the epidemiology of hip fractures. *Clin Orthop.* 1980;152:35–43.
2. Farmer ME, White LR, Brody JA, Bailey KR. Race and sex differences in hip fracture incidence. *Am J Public Health.* 1984;74:1374–1380.
3. Jensen JS. Determining factors for the mortality following hip fractures. *Injury.* 1984;15:411–414.
4. Ions GK, Stevens J. Prediction of survival in patients with femoral neck fractures. *J Bone Joint Surg.* 1987;69-B:384–387.
5. Kenzora JE, McCarthy RE, Lowell JD, Sledge CB. Hip fracture mortality. Relation to age, treatment, preoperative illness, time of surgery, and complications. *Clin Orthop.* 1984;186:45–56.
6. Magaziner J, Simonsick EM, Kashner TM, Hebel JR, Kenzora JE. Survival experience of aged hip fracture patients. *Am J Public Health.* 1989;79:274–278.
7. Fisher ES, Baron JA, Malenka DJ, et al. Hip fracture incidence and mortality in New England. *Epidemiology.* 1991;2:116–122.
8. Fisher ES, Whaley FS, Krushat WM, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health.* 1992;82:243–248.
9. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis.* 1987;40:373–383.
10. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley and Sons; 1981:244–246.
11. *LOGISTIC P. SAS/STAT User's Guide.* Cary, NC: SAS Institute Inc; 1990; 2.
12. *SAS/STAT Software: The PHREG Procedure.* Cary, NC: SAS Institute Inc; 1992.
13. Nilsson LT, Strömqvist B, Thorngren KG. Function after hook-pin fixation of femoral neck fractures. Prospective 2-year follow-up of 191 cases. *Acta Orthop Scand.* 1989;60:573–578.
14. Calandruccio RA. Classification of femoral neck fractures in the elderly as pathologic fractures. Proceedings of the 11th open scientific meeting. *Hip.* 1983;9–33.
15. Hunter GA. Should we abandon primary prosthetic replacement for fresh displaced fractures of the neck of the femur? *Clin Orthop.* 1980;152:158–161.
16. Welch RB. The rationale for primary hemiarthroplasty in the treatment of fractures of the femoral neck in elderly patients. *Hip.* 1983:42–50.
17. Petitti DB, Sidney S. Hip fracture in women. *Clin Orthop.* 1989;246:150–155.
18. Holmberg S, Conradi P, Kalén R, Thorngren KG. Mortality after cervical hip fracture. 3002 patients followed for 6 years. *Acta Orthop Scand.* 1986;57:8–11.
19. White BL, Fisher WD, Laurin CA. Rate of mortality for elderly patients after fracture of the hip in the 1980's. *J Bone Joint Surg [Am].* 1987;69:1335–1340.
20. Beals R. Survival following hip fracture. *J Chron Dis.* 1972;25:235–244.
21. Jacobsen SJ, Goldberg J, Miles TP, Brody JA, Stiers W, Rimm AA. Race and sex differences in mortality following fracture of the hip. *Am J Public Health.* 1992;82:1147–1150.
22. Franks P, Clancy CM, Gold MR. Health insurance and mortality—evidence from a national cohort. *JAMA.* 1993;270:737–741.
23. Magaziner J, Simonsick EM, Kashner TM, Hebel JR, Kenzora JE. Predictors of functional recovery one year following hospital discharge for hip fracture: a prospective study. *J Gerontol.* 1990;45:M101–107.
24. Mossey JM, Mutran E, Knott K, Craik R. Determinants of recovery 12 months after hip fracture: the importance of psychosocial factors. *Am J Public Health.* 1989;79:279–286.
25. Skinner P, Riley D, Ellery J, Beaumont A, Coumine R, Shafighian B. Displaced subcapital fractures of the femur: a prospective randomized comparison of internal fixation, hemiarthroplasty and total hip replacement. *Injury.* 1989;20:291–293.
26. Rodriguez J, Herrara A, Canales V, Serrano S. Epidemiologic factors, morbidity and mortality after femoral neck fractures in the elderly. A comparative study: internal fixation vs hemiarthroplasty. *Acta Orthop Belg.* 1987;53:472–479.
27. Broos PL, Stappaerts KH, Luiten EJ, Gruwez JA. Endoprosthesis. The best way to treat unstable intracapsular hip fractures in elderly patients. *Unfallchirurg.* 1987;90:347–350.
28. Wentworth DN, Neaton JD, Rasmusser WL. An evaluation of the Social Security Administration Master Beneficiary Record File and the National Death Index in the ascertainment of vital status. *Am J Public Health.* 1983;73:1270–1274.

Lumbar Spine Surgery and Mortality among Medicare Beneficiaries, 1986

Neil B. Oldridge, PhD, Zhong Yuan, MD, James E. Stoll, MD, and Alfred R. Rimm, PhD

ABSTRACT

Objectives. The purpose of this study was to compare lumbar spine surgical procedures by age, gender, and number of comorbidities with respect to mortality in patients 65 years of age and older in the United States.

Methods. A 100% sample of the 1986 Medicare inpatient Health Care Financing Administration claims files databases involving lumbar spine surgical procedures was analyzed.

Results. Lumbar spine surgery in 34 418 patients (median age = 71 years) was associated with a significant increase in in-hospital and 1-year cumulative mortality only beyond 80 years of age. When adjusted for age, in-hospital and 1-year cumulative mortality with both decompression and excision procedures were significantly higher in men than in women. When adjusted for both age and gender, mortality increased significantly as the number of comorbidities increased.

Conclusions. With lumbar spine surgery in elderly patients, mortality did not significantly increase until 80 years of age and was consistently associated with decompression and excision, with male gender, and with an increase in number of comorbidities. (*Am J Public Health.* 1994;84: 1292–1298)

Introduction

Data on the effectiveness of lumbar spine surgery in elderly patients (those 65 years of age and older) are rare. We were able to identify few outcome studies that supported or negated Parrish's 1962 recommendation that age, at least up to 72 years, should not deprive patients of the benefits of surgery in the treatment of "lumbar disc syndrome" if the appropriate indications exist.[1] We were also able to identify only one randomized trial of surgery vs conservative care, that carried out by Weber[2] with 126 patients 22 to 55 years of age who had herniated lumbar discs. There was significant improvement at 1 year in those who had undergone surgery; by 4 years, however, the differences between surgery and conservative care were no longer significant. A literature search for low back pain and spine surgery citations in older patients identified nine articles published between 1987 and 1991 involving a total of 370 patients (mean age = 66.6 years; range = 45 to 89 years), approximately equal numbers of whom were male and female.[3–11] In a recent study based on a hospital discharge registry for the state of Washington, mortality, morbidity, and complications following lumbar spine procedures were demonstrated to increase with increasing age, particularly among those patients more than 64 years of age, and with more extensive surgery such as fusion and procedures associated with spinal arthrodesis.[12]

Increasingly, elderly people demand medical care, including surgery, not only to save life but to increase the quality of it. There are a number of treatments for low back pain with leg pain and refractory neurologic deficit, from conservative care to surgery. With the increasing sophistication and range of availabie treatment options, lumbar spine surgery for low back pain in the elderly will no doubt continue to be more prevalent in the future, accelerated by the increasing number of older patients and by the rising demand for surgery by older patients seeking an improved quality of life.

The 1989 National Health Interview Survey,[13] ranked orthopedic impairment as the most frequently reported chronic condition after arthritis and hearing impairment among patients 65 years of age and older. Back impairment, including intervertebral disc disorder, accounted for the largest proportion of orthopedic impairment,[13] and, after emphysema, it accounted for the most restricted activity and bed disability days.[14] Medical advances in anesthesia techniques, systems monitoring, and spine surgery instrumentation have made spinal surgery possible for an increasing number of older patients. However, the sheer size of the increase in surgery between 1979 and 1989 (e.g., a 204% increase in excision, a 232% increase in decompression, and a more than 350% increase in fusion[15,16]) is staggering, especially in light of the pau-

Neil B. Oldridge is with the Department of Health Sciences, University of Wisconsin–Milwaukee, and the Division of Health Services Research, Medical College of Wisconsin, Milwaukee. At the time this paper was written, Zhong Yuan and Alfred R. Rimm were with the Medical College of Wisconsin. They are now with the Department of Epidemiology, Case Western Reserve University Medical School, Cleveland, Ohio. James E. Stoll is with Columbia Hospital, Milwaukee.

Requests for reprints should be sent to Neil B. Oldridge, PhD, Department of Health Sciences, University of Wisconsin–Milwaukee, PO Box 413, Milwaukee, WI 53201.

This paper was accepted February 15, 1994.

city of outcome results and the lack of randomized clinical trials in elderly patients with lumbar spine surgery for low back pain. The striking dichotomy between the relatively stable prevalence of back impairment and the more than twofold increase in lumbar spine surgery rates for patients 65 years old and older deserves close attention.

Development of guidelines for medical and surgical practices through assessment of patient outcomes demands evaluation of the effectiveness of medical and surgical interventions.[17] Access to medical care is becoming a more critical issue in medicine, particularly for elderly patients, who make up the most rapidly increasing population segment in the country.[18] As access becomes more important, analysis of existing national databases (e.g., the Health Care Financing Administration's [HCFA] Medicare databases) should provide useful information for policymakers, practitioners, and patients. In 1986, Medicare covered hospitalization expenses for about 95% of the US population 65 years of age and older.[19] Although patients included in these databases are not randomized to intervention, the databases have been used extensively,[20–28] and recent reports have demonstrated that they provide a valuable research tool.[29–33]

Our primary purpose in this descriptive epidemiological study was to gather data on in-hospital, 30-day postdischarge, and 1-year mortality following lumbar spine surgery in the elderly in the United States. We analyzed a 100% sample of the 1986 Medicare Provider Analysis and Review inpatient HCFA claims files databases for patients 65 years of age and older (eliminating a primary cancer diagnosis) undergoing surgery for low back pain. We then compared surgical procedures by age, gender, and number of comorbidities with respect to mortality.

Methods

The data for this study were derived from HCFA's 1986 Medicare Provider Analysis and Review inpatient claims files records, which include the nature of the patient's medical status and hospital stay from the hospital billing claim form filed with Medicare for the hospitalization (Medicare Part A, UB82 billing forms). The data fields include up to five International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) diagnosis codes; up to three procedure codes; the diagnostic-related group billing category claimed; dates of admission and discharge; and information on summary charges. Additional data are obtained from other HCFA and social security files on the level of Medicare reimbursement; patient (beneficiary) age, race, and sex; zip code of beneficiary's residence; and date of death of beneficiary. Date of death information is updated from the Social Security Master File at 3-month intervals for 2 years following the end of the reporting year (i.e., in this instance, through March 31, 1988). Population and annual mortality statistics for 1986 were obtained from data published by the US Bureau of the Census[34] and the National Center for Health Statistics.[35]

TABLE 1—Gender and Age in Medicare Patients Admitted with Diagnosis and Procedure for Anatomical Low Back Pain, 1986

	No.	Female, %	Median Age, y
First admission			
Spondylosis	74 528	66.6	75
Disc	40 698	63.1	72
Stenosis	68 604	68.1	74
Spondylolisthesis	1 982	75.1	75
Spondylosis plus disc	8 103	66.0	73
Spondylosis plus stenosis	10 311	70.9	75
Disc plus stenosis	11 053	65.3	72
Spondylosis, disc, stenosis	2 286	66.3	73
Total	217 565	66.1	73
First procedure			
Decompression	16 796	56.6	72
Excision	12 312	55.8	70
Fusion	550	70.2	72
Decompression plus excision	3 166	57.0	71
Fusion plus decompression	1 206	72.6	70
Fusion plus excision	388	63.9	70
Lumbar spine surgery	34 418	57.2	71
No surgery	181 204	66.8	73
Surgery for other reasons[a]	1 943	55.0	71

Note. Admissions involving a primary cancer diagnosis were excluded.
[a]Primarily fracture of vertebral column and other, multiple, ill-defined dislocations.

Hospitalization records for all 1986 Medicare patients 65 years of age or older were screened for the initial 1986 admission to the hospital with an anatomical low back pain diagnosis that may or may not have resulted in lumbar spine surgery. ICD-9-CM diagnostic codes for anatomical low back pain, excluding principal cervical and thoracic diagnoses and primary cancer diagnoses (codes 141.0 to 160.9, 162.0 to 172.9, and 174.0 to 209.89), indicated in at least one of the five diagnostic fields of the record included spondylosis (721.0 to 721.99), disc disease (722.0 to 722.99), spinal stenosis (724.0 to 724.99), and spondylolisthesis (738.4). The same hospitalization records were searched separately for an initial lumbar spine surgical procedure in 1986. Selection of lumbar spine surgery cases was based on the first 1986 file record with an ICD-9-CM procedure code indicated in at least one of the three procedure fields of the record: exploration and decompression (laminectomy) of spinal canal structures (03.00, 03.09), excision or destruction (discectomy) of intervertebral disc (80.5, 80.50, 80.51, 80.59), and spinal fusion (at one or more vertebral levels, with and without excision, decompression, or instrumentation) (81.00, 81.06, 81.07, 81.09). The two files were merged, and patients with lumbar spine surgery were grouped by 5-year age increments from 65 to 84 years, with all patients 85 years old and older constituting the last group.

Mortality was defined as occurring before discharge (in-hospital mortality), within 30 days of discharge (30-day postdischarge mortality), or within 1 year of the date of the first hospital admission (1-year mortality) for surgery for low back pain in 1986. Mortality for each procedure (i.e., excision, decompression, decompression plus excision, and fusion, including fusion combinations) and mortality by age group and gender were calculated with the respective number of patients undergoing surgery as the denominator. Fusion, fusion plus decompression, and fusion plus excision were combined in one

TABLE 2—In-Hospital, 30-Day Postdischarge, and 1-Year Mortality Rates (per 100) in Medicare Patients with Lumber Spine Surgery, by Age Group

Procedure	In-Hospital Mortality						30-Day Postdischarge Mortality						1-Year Mortality					
	65–69	70–74	75–79	80–84	85+	Age Adjusted	65–69	70–74	75–79	80–84	85+	Age Adjusted	65–69	70–74	75–79	80–84	85+	Age Adjusted
Decompression			*															
Men	0 13	0 77	1 19	0 83	3 19[a]	0 89**	0 28	0 28	0 53	0 67	1 06	0 39	3 45	4 35	5 94	7 68[c]	12 77[c]	4 93***
Women	0 26	0 40	0 59	0 72	1 04[b]	0 26	0 23	0 33	0 23	0 51	1 38[a]	0 31	2 20	3 53	4 14[b]	3 98	6 92[a]	3 33
Excision																		
Men	0 32	0 63	0 68	2 05	1 72	0 61**	0 12	0 46	0 80	0 34		0 36*	2 39	3 57	5 57	9 59[c]	13 79[a]	3 98***
Women	0 17	0 28	0 53	1 44[b]	2 86[a]	0 39	0 10	0 19	0 23		0 76[a]	0 15	1 85	2 43	2 85	5 90[b]	7 63[a]	2 52
Decompression plus excision																		
Men		0 43	0 87	1 24	4 17[a]	0 46	0 89		0 43			0 43	2 50	3 22	3 48	7 41	8 33	3 38
Women	0 14		0 58	0 64	2 78[a]	0 31			0 58			0 11	1 72	1 05	4 06	5 77[a]	5 57[a]	2 33
Fusion plus fusion combinations																		
Men	0 36			3 45[b]		0 31	0 36		0 88			0 31	2 90	3 85	3 51	10 35[b]		3 62
Women		0 63	0 42	1 19		0 33	0 15	0 42	1 26	1 19		0 46	1 16	1 88	4 18	9 52[b]		2 31

[a] 85+ vs other ages
[b] 80+ vs other ages
[c] 80+ vs 75–79
*vs 70–74
*p < .025 **p < .01 ***p < .001

category for analysis purposes because their numbers alone by 5-year age group were small, making the interpretation of the analysis of procedure by age and gender difficult, and because significant differences in outcome do not differ with widely differing fusion procedures.[16] On the basis of data recorded at the time of the surgery admission, patients were assigned to diagnostic, or comorbidity, categories based on the ICD-9-CM diagnostic code fields. The relationships between the following factors were studied: age (in 5-year age groups starting at 65 years), gender, and comorbid diagnoses. The most frequently reported comorbidities (excluding cancers) were acute ischemic heart disease (ICD-9-CM codes 410 to 411.99), cerebrovascular disease (430 to 438.99), chronic heart failure (428 to 428.99, 402.01, 402.11, 402.91), diabetes (250 to 250.99), hepatic disease (571 to 572.99), hypertension (401 to 405.99), other heart disease (393 to 398, 412 to 414.99), psychiatric disease (290 to 290.99, 294 to 299.99), pulmonary disease (480 to 493.99, 496 to 496.9), and renal disease (582 to 583.99, 585 to 587.99).

Chi-square analyses were used to test for associations between mortality and a given factor or comorbidity. Analyses of variance were used in comparisons of mortality rates by age group and procedure for all patients 65 years of age and older in the United States admitted to the hospital and undergoing spine surgery for low back pain. In appropriate instances, the relative risk (RR, using the ratio of rates) of mortality in one vs another level of patient characteristic is reported. In order to examine the effects of comorbidity on mortality, patients were selected on the basis of 0, 1, 2, and 3 or more comorbid diagnoses. Multiple logistic regression analysis was used to study factors associated with in-hospital, 30-day postdischarge, and 1-year mortality.

Results

Among Medicare patients 65 years of age and older, there were 217,565 first hospital admissions (median age = 73 years) in 1986 involving a primary anatomical diagnosis of low back pain. Of the patients admitted, 144,692 were women (mean age = 75.8 years, SD = 7.3) and 72,873 were men (mean age = 73.9 years, SD = 6.7); the median age was lowest for patients with disc disease (Table 1). Also, 91.1% were White, 5.1% were Black, and 3.8% were of other races.

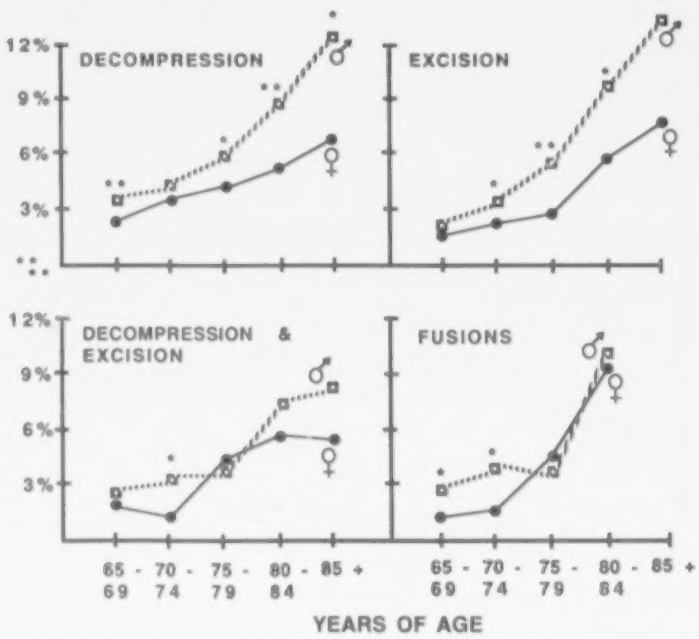
Procedures

There were 34 418 lumbar spine surgical procedures (first procedure in 1986) for low back pain among Medicare patients 65 years of age and older (median age = 71 years), 19 692 involving women (mean age = 72.3 years, SD = 5.4) and 14 726 involving men (mean age = 71.8 years, SD = 5.2). The median age was highest for those undergoing decompression and fusion procedures (Table 1). The proportion of women admitted to the hospital was 66.5%, in comparison with 57.2% of those with a procedure. Of the patients with a procedure, 92.1% were White, 3.8% were Black, and 4.1% were of other races. Patients 80 years of age and older with low back pain accounted for 29% of those who did not have surgery and 10% of those who did. There were 1943 patients who underwent an initial lumbar spine surgery procedure in 1986 but who did not meet the diagnostic eligibility criteria for low back pain; the reasons for surgery in these patients were primarily fracture of the vertebral column, benign neoplasms, and other, multiple, ill-defined dislocations (Table 1). These patients were excluded from the mortality analyses.

The mean length of stay with a procedure was 12.1 days (12.9 days for decompression, 11.4 days for excision, 14.5 days for fusion, 15.0 days for excision plus fusion, 14.0 days for decompression plus fusion, and 12.0 days for decompression plus excision). The estimated hospital charges for 1986 were as follows: $688.57 million for admissions not involving a subsequent surgical procedure (mean length of stay = 7.8 days, mean charge = $3800), $265.02 million for lumbar spine surgical procedures (mean charge = $7700, with fusions the highest at $11 400), and $34.20 million for lumbar spine surgery procedures not associated with a low back pain diagnosis (mean length of stay = 21.0 days, mean charge = $17 600). Thus, the total hospital charges for all 217 565 admissions amounted to an estimated $987.79 million.

Mortality

The in-hospital mortality rate among the 34 418 patients undergoing a first spine surgical procedure for low back pain in 1986 was 0.52% (0.72% in men vs 0.37% in women; $P < .001$). When adjusted for age, in-hospital mortality was higher in men than in women for decompression and for excision procedures ($P < .01$) (Table 2). In-hospital mortality with decompression and decompression plus excision procedures increased significantly at 85 years of age in both men and women, although numbers were small for the combined procedure; with excision, mortality increased significantly at 80



Note. □ = male; ● = female. *$P < .05$, **$P < .005$ (male vs female).

FIGURE 1— One-year cumulative mortality, by procedure and age, in 34 418 Medicare patients undergoing lumbar spine surgery in 1986.

TABLE 3—Relative Risk of 1-Year Mortality Following Lumbar Spine Surgery in Medicare Patients 65 Years of Age and Older in Comparison with Expected Mortality for US Population of the Same Age

Procedure	Age Distribution, y				
	65–69	70–74	75–79	80–84	85+
Decompression					
Men	1.17	0.93	0.85	0.71	0.70
Women	0.32	1.36	1.02	0.58	0.48
Excision					
Men	0.81	0.77	0.79	0.88	0.76
Women	1.11	0.93	0.70	0.86	0.53
Decompression plus excision					
Men	0.84	0.69	0.50	0.68	0.46
Women	1.02	0.40	1.00	0.84	0.81
Fusion plus fusion combinations					
Men	0.98	0.83	0.50	0.96	. . .[a]
Women	0.69	0.72	1.03	1.39	. . .[a]

[a]No deaths.

TABLE 4—Relative Risk (RR) for In-Hospital, 30-Day Postdischarge, and 1-Year Mortality, Adjusted for Age and Gender

Mortality and Comorbidity	RR	95% CI
In-hospital		
1 comorbidity	4.79	3.61, 6.35
2 comorbidities	12.50	9.05, 17.27
3 or more comorbidities	21.59	12.38, 37.64
30-day postdischarge		
1 comorbidity	1.79	1.19, 2.78
2 comorbidities	4.69	2.89, 7.63
3 or more comorbidities	10.49	4.66, 23.59
1-year cumulative		
1 comorbidity	2.18	1.93, 2.45
2 comorbidities	4.03	3.46, 4.78
3 or more comorbidities	8.63	6.47, 12.08

Note. The relative risk for no comorbidity was 1.00. CI = confidence interval.

years of age only in women, and, with fusions, mortality also increased significantly at 80 years of age but only in men (Table 2). The highest in-hospital mortality for men was seen with decompression and decompression plus excision procedures; the highest mortality for women involved excision and decompression plus excision procedures.

The overall 30-day postdischarge mortality rate was 0.31%, (0.38% in men and 0.25% in women; $P < .05$). Age-adjusted, gender-specific 30-day postdischarge mortality was higher in men than women only for excision ($P < .025$) (Table 2). Thirty day postdischarge mortality increased ($P < .05$) at 85 years of age only in women and only with decompression and excision.

The 1-year mortality rate was 3.52% (4.38% in men vs 2.89% in women; $P < .001$). Age-adjusted, gender-specific 1-year mortality was significantly higher in men than women with decompression and excision ($P < .0001$) (Table 2). One-year mortality increased significantly at 80 years of age with excision and fusion procedures in both men and women; the age pattern for the other procedures was not as consistent, but the rates were higher in older subjects. The gender-specific 1-year mortality for each of the procedures is presented graphically in Figure 1; in all instances in which there were significant differences between the genders, men had the higher event rate. The relative risk for 1-year mortality with spinal surgery for low back pain, in comparison with the expected mortality in the US population 65 years old and older by 5-year age groups,[34] was less than 1.00 in more than 75% of the Age × Gender × Procedure cells (Table 3).

The univariate techniques used to identify associations between the various comorbidities identified in the UB82 billing forms and mortality showed significant relative risks of 2.00 and greater for in-hospital, 30-day postdischarge, and 1-year mortality in regard to acute ischemic disease (RRs = 32.9, 5.6, and 7.4, respectively), renal disease (RRs = 9.9, 7.5, and 9.1), other heart disease (RRs = 4.7, 2.0, and 2.0), and cerebrovascular disease (RRs = 3.2, 6.0, and 3.1). Pulmonary disease and congestive heart failure were significant for in-hospital (RRs = 3.8 and 3.3) and 1-year mortality (RRs = 2.1 and 3.0), while hepatic disease was significant for 30-day postdischarge (RR = 27.3) and for 1-year mortality (RR = 7.3). Hypertension was consistently protective (RRs = 0.57, 0.51, and 0.80 for in-hospital, 30-day postdischarge, and 1-year mortality, respectively).

The multiple logistic regression model was then used to determine the relationships between age- and gender-adjusted mortality and number of comorbidities. A relative risk of 1.00 was used for mortality observed in patients with no comorbidities listed; in-hospital, 30-day postdischarge, and 1-year mortality, as expected, increased with the number of comorbidities (Table 4).

Discussion

Although health services data derived from administrative billing information may be criticized when used for clinical decision making,[29,30,33] recent studies demonstrate that hospital billing records can, with careful planning (e.g., consideration of age, gender, and comorbidity) provide a valuable research tool.[29–33] While the accuracy of diagnostic and comorbidity coding remains a concern,[29,30,33] the coding of major surgical procedures in the Medicare claims databases is highly accurate, ranging from sensitivities of .88 to more than .95 for 10 of the 15 procedures examined and positive predictive values of .89 to more than .99 for 7 of the 15 procedures examined.[29] Medicare claims databases similar to those used in this study have been the basis of extensive seminal research in the field of health services.[19–21,23,24] Recently we used the same Medicare databases for reports involving hospital characteristics and mortality rates,[25] hip fracture,[26,27] and factors influencing coronary artery bypass surgery.[28]

In 1986, 217 565 Medicare beneficiaries 65 years of age and older were admitted to US hospitals with a diagnosis of anatomical low back pain and 34 418 underwent lumbar spine surgery. The cumulative 1-year mortality in patients with lumbar spine surgery was 3.52% (of which approximately 15% had occurred before discharge from hospital) and was significantly higher in men than in women. The age-adjusted 1-year mortality ranged from a low of 2.31% in women with fusion and fusion combination procedures to a high of 4.93% in men with decompression procedures. One-year mortality data for patients of comparable age with spinal surgery for low back pain are not available, but the expected US population rate of 3.61% for this age group[30] is similar to that observed here. On the other hand, comparable data in older patients undergoing spinal surgery for low back pain are available for in-hospital mortality rate; this rate was 0.52% in the present study and, again, significantly higher in men than in women, as previously reported by Deyo et al.[12]

The 80- to 85-year age range appears to be a threshold for a significant increase for both in-hospital and 1-year mortality, reinforcing the importance of factors other than age when selecting patients for spinal surgery. Ramirez and Thisted[37] reported an in-hospital mortality rate of 0.059% in patients over the age of 18 years undergoing lumbar discectomy, while Deyo and colleagues[12] reported an overall in-hospital mortality rate of 0.07%, increasing with age from 0.01% at 41 to 64 years, to 0.21% at 65 through 74 years, and 0.60% at 75 years and older. In the present study, the 0.28% in-hospital mortality rate with lumbar spine surgery for low back pain in patients 65 to 74 years of age was similar to that reported by Deyo et al. for patients in the same age range[12]; however, in those patients 75 years of age and older, the rate of 1.3% observed in this study was approximately twice that reported by Deyo et al. for patients of the same age.[12] The large difference between the overall in-hospital mortality reported by Ramirez et al.[37] and that reported here is due, in part, to the differences between

the studies with respect to the procedures considered for analysis and with respect to age (the patients included in the Ramirez study underwent lumbar discectomy and were 18 years of age and older[37]).

A desire for an improved quality of life on the patient's part may be the most important factor driving the increase in number of lumbar spinal procedures. Today the elderly live longer, healthier, and more active lives, and factors such as life expectancy, overall health status, and the ability of patients to tolerate the surgery and subsequent rehabilitation need to be considered when making treatment decisions. The importance of the discrepancy between functional and chronological age, and therefore the need to question age per se as an exclusionary criterion for surgery, has been reinforced in recent investigations of various surgical procedures (e.g., coronary bypass surgery in octogenarians[38] and nonagenarians[39]). Logically, increasing age and associated medical conditions, together with the stress of surgery, should be expected to result in increased mortality. However, appropriately selected patients 90 years of age and older do tolerate the stress of surgery relatively well,[39] although age bias might be more widespread than warranted, at least for elderly patients with either cancer[40,41] or heart disease.[38,42,43] The 1-year mortality rate among the 181 204 patients admitted with low back pain but without subsequent surgery was 10.01%. We can find no publications describing the treatment and outcome of hospitalization in elderly patients who are admitted to the hospital for low back pain but do not undergo lumbar spine surgery. However, the difference in mortality in favor of patients undergoing surgery is probably associated with a patient selection bias inherent in making decisions about older patients as candidates for lumbar spine surgery. When patient age was less than 75 years, the relative risk for 1-year mortality, in comparison with the expected mortality in the US population 65 years old and older,[34] was greater than 1.00 in 25% of the Age × Gender × Procedure cells; however, it was greater than 1.00 in only 12.5% of the cells when patient age was 75 years or older. Patient selection, particularly with increasing age, appears to be as, or more, important than the procedure when considering patient outcome. In the present study, at least up to the age of 80 years, the low mortality associated with lumbar spinal procedures for low back pain extends the relevance of Parrish's 1962 recommendation that, if the appropriate indications exist, age per se should not deprive patients of the benefits of surgery in the treatment of low back pain.[1]

While the age effect per se, at least between 65 and 80 years, in those undergoing lumbar spine surgery does not appear to greatly increase the risk of death after such surgery, the number of comorbid conditions does increase this risk, supporting observations with other surgical procedures in the elderly.[12,38,39] The extent to which comorbid conditions in this study affected mortality was quite marked; the relative risk for in-hospital mortality increased from 1.00 for no coexistent comorbidity to 21.59 with three or more comorbidities; the corresponding 30-day postdischarge risk was 10.49, and the 1-year mortality was 8.63. Factors that probably contribute to the increasing risk (particularly in-hospital and 30-day postdischarge) with lumbar spine surgery seen at 80 years of age and older are an increasingly weaker bone stock and greater neural involvement, an increasingly high likelihood of serious comorbidities, and a poorer physical status (per the American Society of Anesthesiology classification). While the HCFA billing records are not as reliable with respect to comorbidities as they are with respect to surgical procedures,[29,33] the relative risk for in-hospital, 30-day, and 1-year mortality, as might be expected, was significantly increased with one or more comorbidity, particularly with coexistent ischemic heart and renal disease. The paradoxical protective effect of hypertension observed in this study has been reported previously, but without a definitive explanation.[33]

In summary, this study of the large HCFA claims files for 1986 helps define age and gender demographics related to low back pain and outcome following lumbar spine surgery in all elderly Medicare beneficiaries in the United States. In 34 418 elderly patients who underwent lumbar spine surgery, male gender was associated with higher in-hospital and 1-year age-adjusted mortality than that observed in women. Also, 80 years of age appeared to be a threshold for increased mortality, and, as expected, increasing numbers of comorbid conditions were associated with an increased risk of age-adjusted mortality. An important limitation inherent in this Medicare database is the lack of data on the impact of lumbar spinal surgery for low back pain on functional status and the health-related quality of life in the elderly. However, the low mortality associated with lumbar spinal procedures for low back pain, at least in patients up to 80 years of age, extends the relevance of Parrish's[1] aforementioned recommendation. □

Acknowledgments

This study was supported by grant 1-R15-AG09005-01 from the Institute of Aging, National Institutes of Health.

We would like to thank Dr M. Gottlieb for his assistance in the early phases of the study.

References

1. Parrish TF. Lumbar disk surgery in patients over 50 years of age. *South Med J.* 1962;55:667–669.
2. Weber H. Lumbar disc herniation: a controlled prospective study with ten years of observation. *Spine.* 1983;8:131–140.
3. Herkowitz HN, Kurz LT. Degenerative lumbar spondylolisthesis with spinal stenosis. A prospective study comparing decompression with decompression and intertransverse process arthrodesis. *J Bone Jt Surg.* 1991;73A:802–808.
4. Maistrelli GL, Vaughan PA, Evans DC, Barrington TW. Lumbar disc herniation in the elderly. *Spine.* 1987;12:63–66.
5. Avidan D, Schiffer J, Habot B. Neurological approach to the aging patient. *Mt Sinai Med J.* 1987;54:141–143.
6. Lehmann TR, Spratt KF, Tozzi JE, et al. Long-term follow-up of lower lumbar fusion patients. *Spine.* 1987;12:97–104.
7. Herron LD, Trippi AC. L4-5 degenerative spondylolisthesis. The results of treatment by decompressive laminectomy without fusion. *Spine.* 1989;14:534–538.
8. Mauersberger W, Nietgen T. Surgical treatment of lumbar stenosis: long-term results. *Neurosurg Rev.* 1989;12:291–295.
9. Conley FK, Cady CT, Lieberson RE. Decompression of lumbar spinal stenosis and stabilization with Knodt rods in the elderly patient. *Neurosurgery.* 1990;26:758–763.
10. Johnsson K, Uden A, Rosen I. The effect of decompression on the natural course of spinal stenosis. *Spine.* 1991;16:615–619.
11. Katz JN, Lipson SJ, Larson MG, McInnes JM, Fossel AH, Liang MH. The outcome of decompressive laminectomy for degenerative lumbar stenosis. *J Bone Jt Surg.* 1991;73A:809–816.
12. Deyo RA, Cherkin DC, Loeser JD, Bigos SJ, Ciol MA. Morbidity and mortality with operations on the lumbar spine. The influence of age, diagnosis, and procedure. *J Bone Jt Surg.* 1991;74A:536–543.
13. National Center for Health Statistics. Current estimates from the National Health Interview survey, United States, 1989. *Vital Health Stat [10].* 1990;176.
14. National Center for Health Statistics. Disability days, United States, 1983. *Vital Health Stat [10].* 1986;158.
15. *Detailed Diagnoses and Surgical Procedures for Patients Discharged from Short-Stay Hospitals: United States, 1979.* Hyattsville, Md: National Center for Health Statistics; 1982.
16. National Center for Health Statistics. De-

tailed diagnoses and procedures, national hospital discharge survey. *Vital Health Stat [13].* 1991;108.
17. Roper WL, Winkenwerder W, Hackbrath GM, Krakauer H. Effectiveness in health care. *N Engl J Med.* 1988;319:1197–1202.
18. *Projections of the Population of the United States by Age, Sex, and Race.* Washington, DC: US Dept of Commerce, Bureau of the Census; 1989.
19. *Medicare Enrollment, 1986–87.* Baltimore, Md: Health Care Financing Administration; 1989. HCFA publication 03282.
20. Wennberg JE, Freeman JL, Culp WJ. Are hospital services rationed in New Haven or over-utilized in Boston? *Lancet.* 1987;1: 1185–1189.
21. Wennberg JE, Roos N, Schori A, Jaffe R. Use of claims data systems to evaluate health care outcomes: mortality and reoperation following prostatectomy. *JAMA.* 1987;257:933–936.
22. Wennberg JE, Freeman JL, Shelton RM, Bubolz TA. Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. *N Engl J Med.* 1989;321: 1168–1173.
23. Chassin MR, Brook RH, Park RE, et al. Variations in the use of medical and surgical services by the Medicare population. *N Engl J Med.* 1986;314:285–290.
24. Holahan J, Berenson RA, Kachavos PG. Area variations in selected Medicare procedures. *Health Aff.* 1990;166–175.
25. Hartz AH, Krakauer H, Kuhn EM, et al. Hospital characteristics and mortality rates. *N Engl J Med.* 1989;321:1720–1725.
26. Jacobsen SJ, Goldberg J, Miles TP, Brody JA, Stiers W, Rimm AA. Regional variation in the incidence of hip fracture. *JAMA.* 1990;264:500–502.
27. Jacobsen SJ, Goldberg J, Miles TP, Brody JA, Stiers W, Rimm AA. Hip fracture incidence among the old and very old: a population-based study of 745 435 cases. *Am J Public Health.* 1990;80:871–873.
28. Goldberg KC, Hartz AJ, Jacobsen SJ, Krakauer H, Rimm AA. Racial and community factors influencing coronary artery bypass graft surgery for all 1986 Medicare patients. *JAMA.* 1992;267:1473–1477.
29. Fisher ES, Whaley FS, Krushat WM, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health.* 1992;82: 243–248.
30. Steinberg EP, Whittle J, Anderson GF. Impact of claims data research on clinical practice. *Int J Tech Assess Health Care.* 1990;6:282–287.
31. Fisher ES, Baron JA, Malenka DJ, Barrett J, Bubolz TA. Overcoming potential pitfalls in the use of Medicare data for epidemiologic research. *Am J Public Health.* 1990;80:1487–1490.
32. Bright RA, Avorn J, Everitt DE. Medicaid data as a resource for epidemiologic studies: strengths and limitations. *J Clin Epidemiol.* 1989;42:937–945.
33. Iezzoni LI, Foley SM, Daley J, Fisher E, Heeren T. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA.* 1992;267:2197–2203.
34. *Current Population Reports.* Washington, DC: US Dept of Commerce; 1988.
35. *Vital Statistics of the United States 1986, Volume II—Mortality.* Washington, DC: US Dept of Health and Human Services; 1988.
36. Turner JA, Ersek M, Herron L, et al. Patient outcomes after lumbar spinal fusions: a comprehensive literature synthesis. *JAMA.* 1992;268:907–911.
37. Ramirez LF, Thisted R. Complications and demographic characteristics of patients undergoing lumbar discectomy in community hospitals. *Neurosurgery.* 1989;25: 226–231.
38. Mick MJ, Simpfendorfer C, Arnold AZ, Piedmonte M, Lytle BW. Early and late results of coronary angioplasty and bypass in octogenarians. *Am J Cardiol.* 1991;68: 1316–1320.
39. Hosking MP, Warner MA, Lobdell CM, Offord KP, Melton JL III. Outcomes of surgery in patients 90 years of age and older. *JAMA.* 1989;261:1909–1915.
40. Bennett CL, Greenfield S, Aranow H, Ganz P, Vogelzang NJ, Elashoff RM. Patterns of care related to age of men with prostate cancer. *Cancer.* 1991;67:2633–2641.
41. Greenfield S, Blanco DM, Elashoff RM, Ganz PA. Patterns of care related to age of breast cancer patients. *JAMA.* 1987;257: 2766–2770.
42. Montague TJ, Ikuta RM, Wong RY, Bay KS, Teo KK, Davies NJ. Comparison of risk and patterns of practice in patients older and younger than 70 years with acute myocardial infarction in a two-year period (1987–1989). *Am J Cardiol.* 1991;68:843–847.
43. Smith SC, Gilpin E, Ahnve S, et al. Outlook after acute myocardial infarction in the very elderly compared with that in patients aged 65 to 75 years. *J Am Coll Cardiol.* 1990;16:784–792.

Supplement on Exercise and Aging Available

"Exercise and Aging: Prescription for Wellness," a supplement to the May 1994 issue of *Southern Medical Journal*, is available from the Southern Medical Association. The supplement reflects the contents of a symposium held in Gainesville, Fla, in September 1992, for the purpose of updating health professionals on developments in the field of basic and applied exercise science for both healthy and chronically ill elderly persons.

Topics covered in the supplement include the effects of exercise on age and disease, skeletal muscle function, bones, metabolism, arthritis, and the nervous and cardiac systems; respiratory and spinal considerations; and motivating and establishing programs for the elderly.

Contact Lisa Rodgers at the Southern Medical Association, PO Box 190088, Birmingham, AL 35219-0088; tel 1-800-423-4992.

Diet, Indicators of Kidney Disease, and Later Mortality among Older Persons in the NHANES I Epidemiologic Follow-Up Study

Johanna T. Dwyer, DSc, RD, Jennifer H. Madans, PhD, Barry Turnbull, PhD, Joan Cornoni-Huntley, PhD, Connie Dresser, RDPH, LN, Donald F. Everett, MA, and Ronald D. Perrone, MD

ABSTRACT

Objectives. The purpose of this study was to determine whether diet adversely affected survival among 2572 older persons with indicators of kidney disease in a population-based cohort. Average follow-up time for survivors, of whom 1453 (57%) had died at analysis, was 14.5 years.

Methods. Kidney disease indicators were a "yes" response to "Has a doctor ever told you that you have kidney disease or renal stones?" and/or trace or greater amounts of protein in urine. Dietary protein intakes were calculated from 24-hour recalls.

Results. Cox proportional hazards models were used, stratified by sex, with age, body mass index, blood pressure, education, smoking status, total caloric intake, and diabetes mellitus as covariates. Relative risk of total mortality with an additional 15 g of protein per day was 1.25 (95% confidence interval [CI] = 1.09, 1.42) among White men with kidney disease indicators, vs 1.00 (95% CI = 0.95, 1.06) among those without them; relative risks of renal-related mortality were 1.32 (95% CI = 0.97, 1.79) and 0.95 (95% CI = 0.81, 1.11), respectively. No significant differences were found for White women.

Conclusions. Once chronic renal disease is present, diet may be associated with earlier mortality in White males. (*Am J Public Health.* 1994;84: 1299–1303)

Introduction

Chronic kidney disease leads to increased mortality and morbidity. Kidney disease, kidney stones, and biochemical signs of kidney pathology, such as protein in the urine, increase risks of chronic renal insufficiency, an incurable condition that frequently progresses to end-stage renal disease.[1] End-stage renal disease is fatal unless dialysis or transplantation is available. Even with these therapies, premature atherosclerosis, stroke, hypertension, and other complications reduce life expectancy.[2]

High blood pressure does, and high protein intakes may, speed the rate of progression of chronic renal insufficiency,[3] and protein restriction may slow it.[4–10]

We analyzed data from the population-based National Health and Nutrition Examination Survey I (NHANES I)[11,12] and the NHANES I Epidemiologic Follow-Up Study[13–15] to determine whether diet adversely affected survival among Whites aged 65 years or older who had indicators of kidney disease at the time of baseline examination and whether the nature of the relationship differed for those with and without kidney disease. These data sets were collected before any dietary adjustments were made for individuals with early chronic renal insufficiency.[16,17]

Methods

The NHANES I, a probability sample of the civilian noninstitutionalized population of the United States,[11] was conducted from 1971 through 1974 and augmented by an additional national sample in 1974 and 1975.[18] Information on disease status, demographic characteristics, dietary intake, and other covariates from NHANES I was used for the analysis.[15] From 1982 through 1984, with additional waves in 1986 and 1987, the NHANES I Epidemiologic Follow-Up Study recontacted NHANES I participants who were 25 years of age and older at their first interview (or suitable proxies for decedents).[13–15] Vital status through 1987 is known for 96% of the original cohort, and death certificates have been obtained for 96% of the decedents. Average length of follow-up for survivors is 14.5 years.

This study was limited to White persons 65–74 years of age at baseline (1339 men and 1497 women) for whom nutrition information was collected. Whites were studied because of an insufficient number of members of other ethnic groups for ethnicity-specific analyses. Those older than 65 years were studied

Johanna T. Dwyer is with Tufts University Medical School, Boston, Mass; the Frances Stern Nutrition Center, New England Medical Center Hospitals, Boston, Mass; and the US Department of Agriculture, Human Nutrition Research Center on Aging, Tufts University. Jennifer H. Madans is with the Division of Analysis, National Center for Health Statistics, Hyattsville, Md. Barry Turnbull is with Alkermes, Cambridge, Mass. Joan Cornoni-Huntley is with the National Institute on Aging, Bethesda, Md. Connie Dresser is with the Public Health Applications Research Branch, National Cancer Institute, Rockville, Md. Donald F. Everett is with the Collaborative Clinical Research Branch, National Eye Institute, Bethesda, Md. Ronald D. Perrone is with Tufts University Medical School and the Division of Nephrology, New England Medical Center Hospitals.

Requests for reprints should be sent to Johanna T. Dwyer, DSc, RD, Box 783, NEMCH, 750 Washington St, Boston, MA 02111.

This paper was accepted March 16, 1994.

TABLE 1—Study Population by Sex and by Kidney Disease Indicators at Baseline

Kidney Disease Indicator at Baseline	Percentage	
	Men (n = 1231)	Women (n = 1341)
Doctor's statement		
Currently has disease	3.6	4.1
Does not have disease	96.4	95.9
Proteinuria		
None	96.0	97.8
Trace	2.3	1.3
Greater than trace	1.7	0.9
Kidney disease risk case status	7.4	6.3

because kidney disease is more prevalent in such persons. Ten men and 17 women were lost to follow-up, and an additional 98 men and 139 women were eliminated from the analysis because of missing data on covariates, resulting in a sample size of 1231 White men and 1341 White women aged 65–74 years at baseline.

Table 1 describes the study population according to disease status. Because NHANES I did not include a definitive measure of kidney disease, either of two indicators of kidney disease was used to identify cases. Persons who responded "yes" to both of the following questions were counted as case subjects: "Has a doctor ever told you that you have any of the following conditions (kidney disease *or* renal stone)? If so, do you still have it?" Persons with a laboratory determination of protein in the urine were also considered to have kidney disease. Protein in the urine was ascertained from urine specimens provided by participants in the mobile examination center by using a urine dipstick.[17] Persons who had a positive result (trace amount or greater, i.e., at least 15 mg of albumin/100 mL) on the dipstick were defined as having kidney disease. One hundred seventy-five cases of kidney disease (91 men and 84 women) were identified by at least one of these criteria. Only two subjects reported kidney disease and had proteinuria.

We also performed analyses using a definition of kidney disease based solely on proteinuria, because substantial, constant, continuous proteinuria is a sensitive, specific indicator of kidney disease that is strongly associated with both morbidity and mortality.[20,21] Trace or low levels of proteinuria, such as those we included as indicators of kidney disease, are relatively nonspecific and may result from many causes, including fever, concentrated urine samples, benign forms of glomeruloproteinuria,[22,23] infection, tubulointerstitial disease, exercise, and other factors.[24,25] Because the dipstick measure of urinary protein is semiquantitative, the midpoint for the protein range of each category was assigned (e.g., trace = 22.5; 1+ = 65; 2+ = 200; 3+ = 650; and 4 = 1000 mg of protein/dL). We further refined the proteinuria measure by calculating the ratio of urinary protein (mg/dL) to creatinine (mg/dL) in order to rule out the possibility that high numbers of false-positives resulted from concentrated urine. Ratios equal to or greater than 0.2 were used to indicate disease. By this definition, 27 men and 15 women had kidney disease.

Protein and other nutrient intakes were calculated from the 24-hour recall of food intake collected during NHANES I. Those who consumed the highest amounts of protein based on the 24-hour recall also consumed the largest number of servings of animal protein–rich foods as indicated on a nonquantitative food frequency questionnaire that was also collected (data not shown). Protein intakes were highly and significantly ($P < .001$ level or less) intercorrelated with other nutrients of possible interest ($r = 0.839$ with phosphorus, 0.566 with calcium, 0.581 with total fat, 0.629 with saturated fat, and 0.776 with calories). Although we used calories as a covariate in the analysis, it was not possible to definitively ascertain if the observed effects were due to protein alone or to one of these associated constituents.

We obtained mortality outcomes through 1987 using all three waves of follow-up. We obtained death certificates for all but 37 (16 men and 21 women) of the 1453 reported deaths and coded the underlying cause of death from them with the *International Classification of Diseases*, 9th revision (ICD-9), using standard procedures.[26] In this report, mortality was analyzed for all causes, as well as for renal and renal-related diseases (e.g., ICD-9 condition codes 580, 599, 016, 179–189, 250, 401–404, and 441–448). Renal-related condition was defined as any disease of the kidney, plus hypertension and its complications, including hypertensive, atherosclerotic cardiac and vascular disease, because of the well-known associations between kidney disease, hypertension, and premature atherosclerosis.[1] Ninety men and 61 women died of kidney-related causes.

Covariates included age, education, body mass index, blood pressure, protein intake, history of diabetes, smoking status, and total caloric intake measured by using standardized procedures specified in the NHANES manual[19] and described elsewhere.[11,12,18] All of these are associated with mortality and possibly with kidney disease.[1,2,3,27]

Age-adjusted death rates were calculated by using the direct method, with the total NHANES I population as the standard population. Estimates of the risk of death and associated confidence intervals (CIs) were derived from Cox proportional hazards regression models computed with the PHGLM procedure in the Statistical Analysis System.[28] Models were estimated separately for men and women. Interaction terms involving kidney disease status and protein intake were included in the models to test the hypothesis that the effect of protein intake on survival is different among those with kidney disease compared with those without the disease and to estimate risk for those with or without a kidney disease. To assess the effect of the complex survey design on the results, weighted person-time logistic regression analyses were done.[28,29] Previous analyses of the NHANES I Epidemiologic Follow-Up Study data indicated that clustering and stratification have only minimal effects on the estimates.[30,31] Because the results of the weighted analyses were consistent with the unweighted estimates, only the estimates from the unweighted Cox regression models are presented.

Results

Table 2 presents characteristics of the study population at baseline. Male case subjects were similar to noncase subjects except that White men with indicators of kidney disease were more likely to report histories of high blood pressure and diabetes. In women, case subjects were older, were heavier as measured by body mass index, and had lower caloric intakes, higher systolic blood pressures, and lower educational levels than those without kidney disease. Diastolic blood pressure, history of high blood

TABLE 2—Characteristics of White Individuals Aged 65 Years and Older at NHANES I in Study Population

	Men		Women	
	With Kidney Disease (n = 91)	Without Kidney Disease (n = 1140)	With Kidney Disease (n = 84)	Without Kidney Disease (n = 1257)
Mean (SD)				
Age, y	68.8 (2.9)	68.9 (2.9)	69.8 (3.0)	69.0 (2.9)
Height, cm	170.1 (7.1)	171.1 (6.6)	156.9 (6.0)	158.2 (6.2)
Weight, kg	73.2 (13.7)	74.8 (12.4)	67.8 (14.0)	65.8 (13.1)
Body mass index	25.3 (4.5)	25.5 (3.9)	27.6 (5.6)	26.3 (5.0)
Calories	1727.0 (686.9)	1793.0 (703.5)	1185.7 (465.2)	1309.6 (479.2)
Protein intake, g/24 h	65.8 (28.1)	70.8 (30.9)	48.6 (24.6)	52.6 (22.7)
Protein intake, g/24 h/1000 calories	0.039 (0.013)	0.040 (0.011)	0.041 (0.013)	0.041 (0.013)
Systolic blood pressure, mm Hg	150.9 (28.1)	145.8 (23.6)	157.9 (27.3)	151.6 (24.6)
Diastolic blood pressure, mm Hg	87.1 (15.3)	85.2 (12.6)	86.8 (16.0)	85.7 (12.3)
Percentage				
History of diabetes	14.3	8.3	13.1	9.6
History of high blood pressure	40.7	26.9	58.3	44.7
Education (completed grade 12 or higher)	25.3	31.5	19.0	36.6
Current smoker	37.4	31.0	15.5	12.9

TABLE 3—Number of Deaths and Age-Adjusted Death Rates per 1000 Years of Follow-Up, by Cause, Kidney Disease Case Status at Baseline, and Sex

	Death from All Causes		Kidney-Related Deaths [a]	
Case Status	No.	Rate per 1000 y	No.	Rate per 1000 y
Men				
Case	71	95.40	14	18.79
Noncase	751	66.72	76	6.76
Women				
Case	48	53.87	8	8.98
Noncase	583	39.62	53	3.62

[a]Included ICD-9 causes of death as follows (with no exclusions for comorbid conditions): codes 580–583 (nephrotic syndrome, chronic glomerulonephritis, nephrosis, etc.); codes 584–589 (other types of renal disease); codes 590–599 (other diseases of the urinary system); code 016 (tuberculosis of kidney and genitourinary system); codes 179–189 (malignant neoplasms of genitals and urinary organs); code 250 (diabetes mellitus); codes 401–404 (hypertension, hypertensive heart or/and renal disease); and codes 441–448 (other diseases of artenes, arterioles, and capillaries).

pressure, weight, and history of diabetes were also greater among case subjects.

Table 3 describes the cohort's mortality experience. Men with kidney disease were 40% more likely to die during the follow-up period and over two and one-half times more likely to die of renal-related disease than were noncase subjects. Women with kidney disease were also 36% more likely to die during follow-up and two and one-half times as likely to die from renal-related disease as noncase subjects.

Table 4 presents the regression coefficients and standard errors from the Cox proportional hazards models for all-cause and renal-related mortality. Among men, all covariates (except systolic blood pressure and education) were significantly related to all-cause mortality. Only systolic blood pressure and diabetes were associated with renal-related mortality, however. In men, there was a significant interaction between kidney disease status and protein, indicating that the effect of protein intake on mortality varied by disease status at baseline. To illustrate this effect, we calculated the relative risks (RRs) and associated confidence intervals (CIs) for an increase in protein intake of 15 g per day for those with and without kidney disease. The effect for those with kidney disease was calculated by using the Cox model coefficients for protein intake and for the interaction term where protein intake was evaluated at any intake amount and then at that amount plus 15 g. The effect for those without kidney disease is only a function of the coefficient for protein intake because the interaction term drops out of the model for this group.

The relative risk of total mortality associated with a 15-g increase in protein intake daily among men with kidney disease indicators was 1.25 (95% CI = 1.09, 1.42), whereas the risk among the nondiseased was 1.00 (95% CI = 0.95, 1.06). A similar effect was found for renal-related mortality (RR = 1.32 [95% CI = 0.97, 1.79] vs RR = 0.95 [95% CI = 0.81, 1.11], respectively). Models stratified on kidney disease status produced very similar effects of protein intake on mortality. In addition, the stratified models were re-estimated with protein intake categorized into quartiles. These analyses provided no evidence of a nonlinear relationship between intake and mortality among case or noncase subjects.

The results for women were quite different. Regardless of their status with respect to kidney disease indicators, protein intake was not associated with either total (RR = 1.12 [95% CI = 0.92, 1.36] with disease vs RR = 1.04 [95% CI = 0.96, 1.11] without disease) or renal-related mortality (RR = 1.16 [CI = 0.72, 1.86] with disease vs RR = 1.06 [95% CI = 0.85, 1.32] without disease).

Because fat consumption (total or saturated) and protein intake are highly related, measures of fat consumption were included in the model to test whether the observed relationship between protein and mortality was confounded by fat intake. No significant interaction between fat and kidney disease was observed for either sex. However, the effect of protein remained significant for men with kidney disease.

We repeated the analysis using the more stringent urinary protein/urinary creatinine ratio definition of kidney disease. Although the interaction between kidney disease status and protein intake

TABLE 4—Regression Coefficients and Standard Errors from Cox Proportional Hazards Models for All-Cause and Renal-Related Mortality

	All-Cause Mortality		Renal Mortality Related	
	Coefficient	SE	Coefficient	SE
Men				
Age	.0713**	.0124	.0628	.0383
Systolic blood pressure	.0016	.0016	.0097*	.0046
History of diabetes (ever)	.5062**	.1145	.9940**	.2992
History of high blood pressure (ever)	.3733**	.0791	.2809	.2392
Body mass index	−0.0275**	.0099	−0.0067	.0303
Educational attainment (over 12 y)	.1430	.0786	.1041	.2345
Current smoker	.3767**	.0748	−0.654	.2444
Calorie intake 24 h	−0.0002*	.0001	.0002	.0002
Protein intake 24 h	0.0002	.0019	−0.0037	.0056
Kidney disease status	−0.5765	.3235	−0.5233	.8315
Interaction: protein by kidney disease status	.0145**	.0043	.0221*	.0100
Women				
Age	.0823**	.0140	.1248**	.0459
Systolic blood pressure	.0067**	.0017	.0031	.0058
History of diabetes (ever)	.5427**	.1179	1.5571**	.2941
History of high blood pressure (ever)	.2560**	.0856	.2702	.2794
Body mass index	−0.0133	.0085	.0125	.0264
Educational attainment (over 12y)	.1971*	.0881	.2181	.2929
Current smoker	.3397**	.1118	−0.1494	.4383
Calorie intake 24 h	−0.0002	.0001	−0.0001	.0004
Protein intake 24 h	.0023	.0024	.0040	.0075
Kidney disease status	−0.0355	.3567	.5273	.9091
Interaction: protein by kidney disease status	.0053	.0065	.0058	.0156

*$P < .05$; **$P < .01$.

was at the marginally significant level for total mortality for men, the magnitude of the effect of an increase in protein intake found in this model was similar to that found in the model based on the more liberal definition of case status. No change in mortality was associated with protein intake for noncase subjects. Among women, significant interactions were not found for total or renal-related mortality, but the number of cases in these latter models was very small.

Discussion

Persons with kidney disease experienced increased mortality, but the possible effects of diet on this association are not well documented.[1] This analysis found a relationship between protein intake and mortality for White males with kidney disease that remained after accounting for possible confounders, including history of diabetes, high blood pressure, and total or saturated fat intake. There is no reason to suspect recall bias in the reporting of diet by disease status or by disease severity.

The study does have several limitations. Twenty-four–hour recalls are relatively insensitive measures of long-term nutrient intake.[32,33] However, those who consumed the highest amount of protein based on the 24-hour recall also consumed the largest number of servings of animal protein–rich foods on a nonquantitative food frequency questionnaire also collected in NHANES I. Underreporting of intakes is common[34] and was likely in this study because reported energy intakes were very low compared with standards.[35–37]

Possibly those classified as case subjects did not have kidney disease, or true cases were missed. The strictest definition of case status limited the first type of misclassification. Analysis with a definition of renal disease based on urinary protein/creatinine ratios identified a smaller numbers of persons (2.2% of men and 1.2% of women), yet results for men were in the same direction as and similar in magnitude to those found when the definition of case status was based on proteinuria and self-report. Other nonrenal causes of proteinuria could not be ruled out, however, and these remain a potential source of false-positives.

We think it unlikely that we failed to identify many patients with renal disease (false-negatives). Either proteinuria or self-report could result in case status. Even those suffering from undiagnosed disease were identified as case subjects if they exhibited proteinuria. Self-report of disease was more likely to have been subject to overreporting errors because it included renal stones (which are not always indicative of renal disease), and some individuals may have confused urinary tract infections with kidney disease.

As for the gender differences found in renal disease indicators and mortality, the definition of kidney disease (which relied on both self-report and proteinuria) may be a less reliable indicator of kidney disease in women, possibly accounting for the lack of association in women. More women than men reported that a doctor had told them they had kidney disease, whereas only half as many women as men had proteinuria in any amount (Table 1). Women, who are more likely to have urinary tract infections than men, may have confused these infections with kidney disease and overreported disease status. Urinary tract infections may also yield a positive dipstick test for low-level proteinuria, but uncomplicated urinary tract infections rarely give rise to later kidney disease.[24] Possibly kidney disease has a different course in men than in women; women are less likely to progress to end-stage renal disease than men, perhaps because they have lower morbidity rates of kidney disease–associated conditions such as hypertension, atherosclerosis, and arteriosclerosis.[1]

Conclusion

If renal disease indicators are present, dietary factors may play a role in heightening mortality in White males 65 years and older. Our study's statistical power and precision were limited due to the small number of persons with kidney disease, and it could not adequately control for comorbidities and unmeasured confounders. Additional analyses are needed to confirm these findings. □

Acknowledgments

This project was funded in part by the NHANES I Epidemiologic Follow-Up Study developed by the National Center for Health Statistics of the Centers for Disease Control, the National Institute on Aging, and other institutes of the National Institutes of Health;

the Charles A. Dana Foundation; and the US Department of Agriculture, Agricultural Research Service, under contract number 53-3K06-5-10.

We thank Mary Beth Ofstedal for statistical assistance and Begabati Lennihan for typing the manuscript.

References

1. Klahr S, Shreiner G, Ichikawa I. The progression of renal disease. *N Engl J Med.* 1988;318:1657–1666.
2. National Institutes of Health. Table 8. Prevalence of end-stage renal disease. *National Kidney and Urologic Diseases Advisory Board 1990 Long Range Plan.* Bethesda, Md: US Dept of Health and Human Services; March 1990:27. NIH publication 90-583.
3. Klahr S. Sounding board: the modification of diet in renal disease. *N Engl J Med.* 1989;320:864–866.
4. Mitch WE. The influence of diet on the progression of renal insufficiency. *Annu Rev Med.* 1989;35:249–264.
5. Brenner BM, Meyer TW, Hostetter TH. Dietary protein intake and the progressive nature of kidney disease: the role of hemodynamically mediated glomerular injury in the pathogenesis of progressive glomerular sclerosis in aging, renal ablation, and intrinsic renal disease. *N Engl J Med.* 1982;307:652–659.
6. Ihle BU, Becker GJ, Whitworth JA, Charlwood RA, Kincaid-Smith PS. The effect of protein restriction on the progression of renal insufficiency. *N Engl J Med.* 1989;321: 1773–1777.
7. Maschio G, Oldrizzi L, Tessitore N, et al. Effects of dietary protein and phosphorus restriction on the progression of early renal failure. *Kidney Int.* 1982;22:371–376.
8. Mitch WE, Walser M, Steinman TI, Hill S, Zeger S, Tungsanga K. The effect of a keto acid–amino acid supplement to a restricted diet on the progression of renal insufficiency. *N Engl J Med.* 1984;35:249–264.
9. Rossman JB, Meijer S, Sluiter WJ, Ter Wee PM, Piers Becht TPM, Donker AJM. Prospective, randomized trial of early dietary protein restriction in chronic renal failure. *Lancet.* 1984;2:1291–1295.
10. Locatelli F, Aberti D, Grazian G, Buccianti G, Redaelli B, Giangrande A. The Northern Italian Cooperative Study Group. Prospective randomized multicenter trial of effect of protein restriction on progression of chronic renal insufficiency. *Lancet.* 1991;337(87 753):1299–1303.
11. National Center for Health Statistics. Plan and operation of the National Health and Nutrition Examination Survey, United States, 1971–73. *Vital Health Stat [1].* 1973;10a. DHEW publication PHS 79–1310.
12. National Center for Health Statistics. Plan and operation of the National Health and Nutrition Examination Survey, United States 1971–73. *Vital Health Stat [1].* 1973;10b. DHEW publication PHS 79–1310.
13. National Center for Health Statistics. Plan and operation of the NHANES I Epidemiologic Follow-up Study 1982–84. *Vital Health Stat [1].* 1987;22. DHHS publication PHS 87–1324.
14. National Center for Health Statistics. Plan and operation of the NHANES I Epidemiologic Follow-up Study, 1986. *Vital Health Stat [1].* 1990;25. DHHS publication PHS 90–1307.
15. National Center for Health Statistics. Plan and operation of the NHANES I Epidemiological Follow-up Study, 1987. *Vital Health Stat [1].* 1992;27. DHHS publication PHS 92–1303.
16. Wohl MB, Goodhart RS. *Modern Nutrition in Health and Disease.* Philadelphia, Pa: Lea and Febiger; 1970.
17. Relman AS, Levinsky NG. Clinical examination of renal function. In: Straus MB, Welt LG, eds. *Diseases of the Kidney.* Boston, Mass: Little, Brown and Company; 1970:87–137.
18. National Center for Health Statistics. Plan and operation of the HANES I Augmentation Survey of Adults 25–74 Years: United States, 1974–75. *Vital Health Stat [1].* 1978;14. DHHS publication PHS 78–1314.
19. *National Health and Nutrition Examination Staff Procedures Manual for the NHANES Survey 1971–73.* Washington, DC: National Center for Health Statistics; 1972: part 15a.
20. Daley RM. Albuminuria and insurability. *Urol Cutan Rev.* 1942;46:13–15.
21. Abuelo JG. Proteinuria: diagnostic principles and procedures. *Ann Intern Med.* 1983;98:186–191.
22. Robinson RR. Isolated proteinuria in asymptomatic patients. *Kidney Int.* 1980;18: 395–406.
23. Yoshimoto M, Tsukahara H, Saito M, et al. Evaluation of variability of proteinuria indices. *Pediatr Nephrol.* 1990;4:136–139.
24. Gapewski J, Gonzales CI, Rich M. Genitourinary disease. In: Singer RB, Levinson L, eds. *Medical Risks.* Lexington, Mass: DC Health Company; 1976:151–156.
25. Sinniah R, Law CH, Pivee HS. Glomerular lesions in patients with asymptomatic persistent and orthostatic proteinuria discovered on routine medical examination. *Clin Nephrol.* 1977;7:1–14.
26. Kleinman JC, Donahue RP, Harris MI, Finucane FF, Madans JH, Brock DB. Mortality among diabetics in a national sample. *Am J Epidemiol.* 1988;128:389–401.
27. Madans JH, Cox CS, Kleinman JC, et al. 10 years after NHANES I: mortality experience at initial followup, 1982–84. *Public Health Rep.* 1986;101:474–478.
28. Hastings RP, ed. *SUGI Supplemental Library User's Guide, Version 5 Edition.* Cary, NC: SAS Institute Inc; 1986:269–293, 436–474.
29. Ingram DD, Kleinman JC. Empirical comparisons for proportional hazards and logistic regression models. *Stat Med.* 1989;8: 526–538.
30. Makuc DM, Kleinman JC. Survival analysis using complex survey data: examples from the NHANES I Epidemiologic Followup Study. Presented at the annual meeting of the American Statistical Association; 1986; Chicago, Ill.
31. Korn EL, Graubard BI. Epidemiologic studies utilizing surveys: accounting for the sample design. *Am J Public Health.* 1991;81: 1166–1173.
32. Loken EB. Twenty-four hour recall. In: Cameron ME, Van Staveren WA, eds. *Manual on Methodology for Food Consumption Studies.* New York, NY: Oxford University Press; 1988:83–87.
33. Bingham SA. The dietary assessment of individuals: methods, accuracy, new techniques, and recommendations. *Nutr Abstracts Rev.* 1987;57(series A):705–742.
34. Schoeller DA. How accurate is self-reported dietary energy intake? *Nutr Rev.* 1990;48:373–379.
35. *Dietary Intake Source Data, United States, 1976–80. Data from the National Health Survey, series II, No. 23.* Hyattsville, Md: National Center for Health Statistics; March 1983:151. DHHS publication PHS 831–1681.
36. Committee on Dietary Allowances, Food and Nutrition Board. *Recommended Dietary Allowances.* 10th ed. Washington, DC: National Academy Press; 1989.
37. Food and Agricultural Organization, World Health Organization. *Energy and Protein Requirements.* Report of a joint FAO/WHO/UNU meeting. Geneva, Switzerland: World Health Organization; 1985.

Emotional Disability Days: Prevalence and Predictors

ABSTRACT

This study considered days missed from work or usual activities for emotional reasons associated with a range of specific psychopathologic disorders, psychosocial distress, and persons found to be asymptomatic. Analyses were performed with the presence or absence of emotional disability days as the dependent variable using logistic regression. The effects of specific mental disorders were compared with the effects of chronic physical conditions for labor force participants and for the total population. The odds ratio (and 95% confidence interval) for subjects with major depressive disorder was 27.8 (6.93, 108.96); for panic disorder, 21.1 (2.25, 198.44); and for schizophrenia, 17.8 (1.73, 182.99). Workplace adjustments for persons with psychopathology are encouraged. (*Am J Public Health*. 1994;84:1304–1307)

Anthony C. Kouzis, PhD, and William W. Eaton, PhD

Introduction

Beyond the immediate tragedy of emotional or mental illness is associated impairment as measured by days lost from work, disability, a reduced sense of well-being, and the need for health care. Functional status—that is, the capacity to perform tasks and activities—is a primary concern of patients, their families, and physicians. Restricted activity days experienced by persons in a year is an important measure of functioning status and well-being.[1-4]

Previous studies have focused on estimated associations of various measures of functioning with specific chronic medical conditions; few associations have been made with patients experiencing distress or psychopathology.[5] Wells et al.[6] and Broadhead et al.[7] have shown that depression and disability from work are related. A positive relationship has been found between physical illness and depression or psychological distress in some[8-14] but not all studies.[15,16] Thus, depression might be related to disability because it is associated with nonpsychiatric medical conditions.

In this research report, we consider days missed from work or usual activities for emotional reasons ("emotional disability days") associated with a range of specific psychopathological disorders and psychosocial distress. We assess and adjust for the influence of sociodemographic characteristics on these relationships, and we compare subjects who are in the labor force full time with the total population.

Methods

Study Sample

The study sample was drawn from the Eastern Baltimore Mental Health Survey conducted during 1981 as part of the National Institute of Mental Health (NIMH) Epidemiologic Catchment Area Program.[17-20] It has recently been reported that East Baltimore has the highest rates of premature disability and death in the state.[21] This paper reports analyses from The Johns Hopkins University site of the NIMH program, which included three contiguous mental health catchment areas with an adult population of approximately 175 000. The data include information on 3481 individuals originally identified through a probability sample of adult household residents. The response rate was 78%.[22] Details concerning the characteristics of this population have been presented elsewhere.[17,18]

Measurement

Each participant in the survey underwent a 90-minute interview that included the NIMH Diagnostic Interview Schedule (DIS), which was scored to reflect diagnoses from the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edition (DSM-III) (hence the label "DIS/DSM-III disorder").[23] Symptoms were counted only if they met the severity criteria of the DIS and were not explained by physical illness, medications, alcohol, or other drug use. This report focuses on disorders reported as present within the 6 months prior to interview.

The authors are with the Department of Mental Hygiene, Johns Hopkins University School of Hygiene and Public Health, Baltimore, Md.

Requests for reprints should be sent to Anthony C. Kouzis, PhD, Johns Hopkins University School of Hygiene and Public Health, 624 N Broadway, Baltimore, MD 21205.

This paper was accepted November 12, 1993.

The Baltimore site included the General Health Questionnaire to measure distress that might not be connected to a specific mental disorder.[24] The standard threshold of four or more was used to define those in the sample considered to be at "high distress."

Specific physical health conditions included asthma, high sugar or diabetes, heart trouble, high blood pressure, arthritis or rheumatism, and trouble breathing. Respondents were also asked if they had ever had a stroke and/or cancer.[25]

Respondents in the interview were asked, "During the last three months, were there any times when you were kept from your work, school or usual activities for at least one whole day because of an injury or because you weren't feeling well?" Those responding affirmatively were asked, "Were you kept from your usual activities because of an illness or physical condition? an emotional problem or trouble with your nerves?" For this study, days missed from work because of emotional problems and physical illness were analyzed separately. As in the Health Interview surveys, 3 months was selected as our recall period, a compromise offering an adequate number of responses and necessary recency.

Full-time employment status was ascertained by asking respondents what they were doing during most of the previous week (as in the Labor Force surveys). Analyses reported in the tables below focus on the 1463 respondents in the labor force at the time. Weighted data are used in our analyses and reported in the text and tables owing to the multistage sampling procedures of the Epidemiologic Catchment Area Program.[26,27]

Results

There is no striking variation according to sociodemographic variables in the percentage of respondents with emotional disability days (Table 1); however, it is notable that those with the least education and those without private insurance have the most absences from work—about 5%. After adjusting for all the personal characteristics, the odds of having an emotional disability day are presented in the far right column of Table 1. However, these were not significant. Household income was entered in regression analyses, but it is not included in the model in Table 2 because it was not significant and because more than 10% of the respondents failed to answer this item.

TABLE 1—Predictors of Emotional Disability Days: Estimates for the Labor Force Population of East Baltimore

	Sample, % (n = 1463)[a]	% Having Emotional Disability Days	Adjusted Odds of Having Emotional Disability Days	95% Confidence Interval
Sex				
Male	49	3.0	1.00	
Female	51	4.5	1.54	0.66, 3.60
Age, y				
18–29	35	2.8	1.00	
30–44	35	4.2	1.49	0.49, 4.50
45–64	26	4.0	1.23	0.38, 3.99
65+	4	4.2	1.20	0.08, 18.91
Education				
Beyond high school	28	2.3	1.00	
High school	36	3.0	1.23	0.35, 4.32
9th to 11th grade	23	5.2	2.14	0.61, 7.51
8th grade or less	13	5.1	2.11	0.48, 9.35
Race				
White and other	67	3.6	1.00	
African American	33	3.8	0.88	0.30, 2.58
Marital status				
Not married	59	3.9	1.00	
Married	41	3.3	0.84	0.32, 2.04
Health insurance				
No insurance	27	4.9	1.00	
Insurance	73	3.2	0.72	0.29, 1.79

[a]Data were required to be complete for age, sex, and race. For other variables, the total does not always equal 1463 owing to missing data.

TABLE 2—Disability Days for Emotional Reasons, by Psychopathology: Estimates for the Labor Force Population of East Baltimore

Type of Psychopathology[a]	Percentage with 1+ Disability Days[b]	No. in Sample with Psychopathology	Adjusted Odds Ratios[c]	95% Confidence Interval
DIS/DMS-III				
Major depressive disorder	44	34	27.8	6.93, 108.96
Alcohol abuse/dependence	16	82	10.9	3.42, 34.91
Drug abuse/dependence	8	36	2.5	0.38, 16.80
Panic disorder	44	10	21.1	2.25, 198.44
Schizophrenia	40	11	17.8	1.73, 182.99
Obsessive-compulsive disorder	11	25	3.5	0.48, 25.98
Phobic disorder	8	174	2.5	0.86, 7.14
Any DIS/DSM-III disorder	11	312	6.6	2.71, 16.28
High General Health Questionnaire score[d]	15	178	8.2	3.32, 20.01

Note. DIS/DSM-III = Diagnostic Interview Schedule/*Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed.
[a]In prior 6 months.
[b]In prior 3 months.
[c]Adjusted for age, education, marital status, race, sex, and health insurance.
[d]In prior several weeks.

Table 2 presents the percentage of those with at least 1 disability day for emotional or mental reasons by type of psychopathology, and the association of psychopathology with emotional disability. The comparison groups for the spe-

cific psychiatric disorders were those not meeting criteria for that specific disorder only. Persons with no DIS/DSM-III disorder were the comparison group for those with any DIS/DSM-III disorder.[28]

The most notable finding is the large percentage of subjects with DIS/DSM-III depressive disorders and panic disorders who reported a disability day owing to emotional reasons (44% each). Closely following are those with schizophrenia (40%). In contrast, among those who did not meet the criteria for any DIS disorder, only 2% reported an emotional disability day.

As the number of disability days was highly skewed, analyses were performed using logistic regression, with the presence or absence of disability days as the dependent variable.[29,30] The association of type of psychopathology with emotional disability is also shown in Table 2. Subjects suffering from major depression had the greatest odds of missing work (odds ratio [OR] = 27.8; 95% confidence interval [CI] = 6.93, 108.96). This was closely followed by those with a panic disorder (OR = 21.1; 95% CI = 2.25, 198.44). Schizophrenics had 18 times the odds of missing work for an emotional problem (OR = 17.8; 95% CI = 1.73, 182.99) compared with those without such a disorder. Those with an alcohol diagnosis were 11 times as likely to miss work at least 1 day (OR = 10.9; 95% CI = 3.42, 34.91). The crude odds ratios differ trivially. The large confidence intervals caution us regarding the possible stability of these findings.

The mean days absent from work for an emotional reason range from 3.2 to 9.4. Depression is associated with the greatest number of absences ($\bar{X}$ = 9.4 days, SD = 6.74), followed by high distress ($\bar{X}$ = 7 days, SD = 6.40). The mean for depression is greater than that owing to all conditions except cancer and cardiovascular problems.

The association of psychopathological disorders with absences from work owing to emotional reasons for the entire population was also examined. The odds ratios for those with panic disorders were 27 times greater, followed by only 9 times greater odds for major depressives. The pattern of other results was similar for the labor force group and the total population.

Discussion

Mental disorder has a strong association with missing days from work among those in the labor force. Subjects experiencing a major depressive disorder had the strongest association with disability as they showed a 27 times greater likelihood of being disabled from participating in work activities. Subjects with a diagnosis of panic disorder followed closely. The effects of sociodemographic characteristics are not critical to an understanding of the determinants of work absences.

Nagi differentiated the vocationally disabled from the vocationally nondisabled in that the latter were able to make job adjustments either on their own or with their employers.[31] Given the high prevalence of emotional disorders, greater concern, a more "reasonable accommodation,"[32,33] and continuous supportive employment services may be warranted.[9] Rising costs of services for morbidity and associated disability have focused attention on disabilities and the quality of life at the middle and later stages of life.[34,35] Our results point to the possibility of making adjustments for the emotionally handicapped as is currently done for the physically handicapped.[36,37] □

Acknowledgments

This analysis was supported by NIMH grant MH 47447 and MH 14592.

The ECA Program is a series of five epidemiological research studies performed by independent research teams in collaboration with staff of the division of Biometry and Epidemiology of the National Institute of Mental Health (NIMH). During the period of data collection, the program was supported by cooperative agreements. The NIMH principal collaborators were Darrel A. Regier, MD, Ben Z. Locke, MSPH, William W. Eaton, PhD (October 1, 1978, through October 1, 1983), and Jack Burke, MD (October 1, 1983, through March 1, 1987). The NIMH project officers were Carl A. Taube, PhD, and William Huber. The principal investigators and co-investigators from the five sites were Jerome K. Myers, PhD, Myrna M. Weissman, PhD, and Gary L. Tischler, MD, of Yale University, New Haven, Conn (U01 MH 33870); Morton Kramer, ScD, Ernest Gruenberg, MD, DPH, and Sam Shapiro, MS, of The Johns Hopkins University, Baltimore, Md (U01 MH 33870); Lee N. Robins, PhD, and John Helzer, MD, of Washington University, St. Louis, Mo (U01 MH 33883); Dan Blazer, MD, and Linda George, PhD, of Duke University, Durham, NC (U01 MH 35386); and Marvin Karno, MD, Richard L. Hough, PhD, Javier I. Escobar, MD, M. Audrey Burnam, PhD, and Diane Timbers, PhD, of UCLA (U01 MH 35865).

We thank the editor and two anonymous reviewers for their criticisms.

References

1. Scholes D, LaCroix AZ, Wagner EH, Grothaus LC, Hecht JA. Tracking progress toward national health objectives in the elderly: what do restricted activity days signify? *Am J Public Health.* 1991;81:485–488.
2. Kosorok MR, Omenn GS, Diehr P, Koepsell TD, Patrick DL. Restricted activity days among older adults. *Am J Public Health.* 1992;82:1263–1267.
3. Patrick DL, Stein J, Porta M, Porter CQ, Ricketts TC. Poverty, health services, and health status in rural America. *Milbank Q.* 1988;66:105–136.
4. Golden PM. *Health, United States, 1986* (prevention profile). Washington, DC: US Dept of Health and Human Services; 1986. DHHS publication PHS 87-1232.
5. Mintz J, Mintz LI, Arruda MJ, Hwang SS. Treatment of depression and the functional capacity to work. *Arch Gen Psychiatry.* 1992;49:761–768.
6. Wells KB, Golding JM, Burnam MA. Psychiatric disorder and limitations in physical functioning in a sample of the Los Angeles general population. *Am J Psychiatry.* 1988;145:712–717.
7. Broadhead WE, Blazer DG, George LK, Tse CK. Depression, disability days, and days lost from work in a prospective epidemiologic survey. *JAMA.* 1990;264:2524–2528.
8. Kathol RG, Petty F. Relationship of depression to medical illness: a critical review. *J Affect Dis.* 1981;3:111–121.
9. Langer TS, Michael ST. *Life Stress and Mental Health.* London, England: Free Press of Glencoe; 1963;2.
10. Neff JA, Husaini BA, McCorkel J. Psychiatric and medical problems in rural communities. *Soc Sci Med.* 1980;14A:331–336.
11. Craig TJ, Van Natta P. Disability and depressive symptoms in two communities. *Am J Psychiatry.* 1983;140:598–601.
12. Aneshensel CS, Frerichs RR, Huba GJ. Depression and physical illness: a multiwave, nonrecursive causal model. *J Health Soc Behav.* 1984;25:350–371.
13. Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic conditions: results from the Medical Outcomes Study. *JAMA.* 1989;262:907–913.
14. Wells KB, Stewart AL, Hays RD, et al. The functioning and well-being of depressed patients: results from the Medical Outcomes Study. *JAMA.* 1989;262:914–919.
15. Cassileth BR, Lusk EJ, Strouse TB, et al. Psychosocial status in chronic illness: a comparative analysis of six diagnostic groups. *N Engl J Med.* 1984;311:506–511.
16. Mason JH, Weener J, Gertman PM, Meenan RF. Health status in chronic disease: a comparative study of rheumatoid arthritis. *J Rheumatol.* 1983;10:763–768.
17. Eaton WW, Regier DA, Locke BZ, Taube CA. The Epidemiologic Catchment Area Program in the National Institute of Mental Health. *Public Health Rep.* 1981;96:319–325.
18. Eaton WW, Holzer CE, Von Korff M, et al. The design of the Epidemiologic Catchment Area surveys. *Arch Gen Psychiatry.* 1984;41:942–948.
19. Eaton WW, Kessler LG, eds. *Epidemiologic Field Methods in Psychiatry: The NIMH Epidemiologic Catchment Area Program.* Orlando, Fla: Academic Press; 1985.
20. Robins LN, Regier DA, ed. *Psychiatric Disorders in America: The Epidemiologic*

Catchment Area Study. New York, NY: Free Press; 1991.
21. Baltimore receives $2 million grant for community-based public health program. *Public Health Newsletter.* October 6, 1992:1.
22. Von Korff M, Cottler L, George LK, Eaton WW, Leaf PJ, Burnam A. Nonresponse and nonresponse bias in the ECA surveys. In: Eaton WW, Kessler LG, eds. *Epidemiologic Field Methods in Psychiatry: The NIMH Epidemiologic Catchment Area Program.* Orlando, Fla: Academic Press; 1985:85–98.
23. Robins LN, Helzer JE, Croughan J, Ratcliff KS. The National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Arch Gen Psychiatry.* 1981;38:381–389.
24. Goldberg D. *The Detection of Psychiatric Illness by Questionnaire.* London, England: Oxford University Press; 1972.
25. Kramer M, Simonsick E, Lima B, Levav I. The epidemiological basis for mental health care in primary health care: a case for action. In: Cooper B, Eastwood R, eds. *Primary Health Care and Psychiatric Epidemiology.* London, England: Routledge; 1981:69–96.
26. Leaf PJ, Myers JK, McEvoy LT. Procedures used in the Epidemiologic Catchment Area Study. In: Robins LN, Regier DA, eds. *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study.* New York, NY: Free Press; 1991:11–32.
27. Kessler LG, Felson R, Royall R, et al. Parameter and variance estimation. In: Eaton WW, Kessler LG, eds. *Epidemiologic Field Methods in Psychiatry: The NIMH Epidemiologic Catchment Area Program.* Orlando, Fla: Academic Press; 1985:327–349.
28. Boyd JH, Burke JD, Gruenberg E, et al. Exclusion criteria of DSM-III: a study of co-occurrence of hierarchy-free syndromes. *Arch Gen Psychiatry.* 1984;41:983–989.
29. Fleiss JL, Williams JB, Dubro AF. The logistic regression analysis of psychiatric data. *J Psychiatric Res.* 1986;20:145–209.
30. Cleary PD, Angel R. The analysis of relationships involving dichotomous dependent variables. *J Health Soc Behav.* 1984;25:334–348.
31. Nagi SZ. Disability concepts revisited: implications for prevention. Appendix A. In: Pope AM, Tarlov AR, eds. *Disability in America: Toward a National Agenda for Prevention.* Washington, DC: National Academy Press; 1991:309–327.
32. Parrish J. Reasonable accommodations for people with psychiatric disabilities. Boston, Mass: Boston University Psychiatric Rehabilitation Center; 1991.
33. Kaufman CL. Reasonable accommodation to mental health disabilities at work: legal constructs and practical applications. Presented at the American Sociological Association annual meeting; October 1992; Pittsburgh, Pa.
34. Verbrugge LM. Physical and social disability in adults. In: Hibbard H, Nutting PA, Grady ML, eds. *Primary Care Research: Theory and Methods Conference Proceedings.* Rockville, Md: US Public Health Service, Agency for Health Care Policy and Research; 1991;31–53.
35. Yelin EH. *Disability and the Displaced Worker.* New Brunswick, NJ: Rutgers University Press; 1992.
36. Bowman T. Breaking down barriers: disabled attain rights to access but action lags. *Baltimore Sun.* July 26, 1992:10A.
37. Pear R. US to consider denial of benefits to many disabled: Reagan policy reversed. *New York Times.* April 19, 1992:A1, A20.

The Prevalence and Demographic Predictors of Illicit and Licit Drug Use among Lesbians and Gay Men

William F. Skinner, PhD

ABSTRACT

Studies on illicit and licit drug use among homosexuals of both sexes have focused primarily on gay men, used limited drug measures, and been conducted in cities known for large homosexual populations. This paper examines (1) the prevalence of 12 illicit and licit drugs by sex and age group and (2) the demographic predictors of past-year frequency of marijuana, alcohol, and cigarette use. Organizational mailing lists were used to collect self-report data on 455 homosexuals living in a southern state. Differences were found between gay men and lesbians in the use of specific substances and in the demographic predictors of drug use. (*Am J Public Health.* 1994; 84:1307–1310)

Introduction

Early research on illicit and licit drug use by homosexuals has primarily examined the prevalence of alcohol use and abuse using small, opportunistic samples.[1–13] Only within the last decade have studies begun to report both drug and alcohol use among larger and more representative samples.[14–19] However, most of this research has excluded lesbians, employed limited drug use measures, and been conducted in large cities known for their large homosexual populations (e.g., San Francisco). Little is known about the full extent of illicit and licit drug use among gay men and lesbians living in smaller cities.

This paper presents selected data from the Trilogy Project, a study of self-reported illicit and licit drug use among homosexuals of both sexes from two metropolitan areas in a southern state. Two issues are examined: (1) lifetime, past-year, and past-month age-specific prevalence of use of six illicit and two licit "recreational" drugs (alcohol and cigarettes) as well as the nonmedical use of four psychotherapeutic drugs by age; and (2) demographic predictors of reported frequency of use over the past year of marijuana, alcohol, and cigarettes—the three drugs most commonly used by both lesbians and gay men. Further data are reported elsewhere.[20]

Methods

Sample and Procedures

Respondents were self-defined homosexuals living in and around the two

William F. Skinner is with the Department of Sociology, University of Kentucky, Lexington.

Requests for reprints should be sent to William F. Skinner, PhD, Department of Sociology, University of Kentucky, Lexington, KY 40506-0027.

This paper was accepted November 4, 1993.

Note. The content, interpretations, and conclusions expressed in this paper are exclusively those of the author.

TABLE 1—Lifetime, Past-Year, and Past-Month Prevalence (Percentages) and Numbers of Self-Reported Illicit Drug Use, Non-Medical[a] Use of Psychotherapeutics, and Alcohol and Cigarette Use among Gay Men and Lesbians, by Age

Age, y	Life %	Year %	Month %	No.	Life %	Year %	Month %	No.	Life %	Year %	Month %	No.	Life %	Year %	Month %	No.	Life %	Year %	Month %	No.
	Marijuana				Cocaine				Crack				Inhalants[b]				Hallucinogens[c]			
Gay men																				
18–25	79.2	50.0	37.5	24	26.1	13.0	4.3	23	8.7	0.0	0.0	23	62.5	41.7*	16.7	24	50.2	20.8	8.3	24
26–34	75.5	35.8	16.0	106	25.2	8.4	1.9	107	3.7	0.9	0.0	107	63.2	25.5*	14.2*	106	19.6	2.8	0.0	107
35+	65.4	27.7	15.4	130	27.9	6.2	2.3	129	2.3	1.6	0.0	129	58.8*	26.0*	18.3*	131	27.1	3.1	8.0	129
Total	70.8	33.1	17.7	260	26.6	7.7	2.3	259	3.5	1.2	0.0	259	60.9*	27.2*	16.5*	261	26.2	4.6	1.2	260
Lesbians																				
18–25	82.4	41.2	23.5	17	23.5	11.8	0.0	17	0.0	0.0	0.0	17	41.2	11.8	5.9	17	41.2	11.8	0.0	17
26–34	95.5*	40.9	21.2	66	42.4*	9.1	4.5	66	6.1	0.0	0.0	66	50.8	9.2	3.1	65	45.5*	4.5	4.5*	66
35+	82.7*	22.1	13.5	104	28.8	4.8	1.9	104	1.0	0.0	0.0	103	28.6	2.9	2.9	105	32.7	1.0	0.0	104
Total	87.2*	30.5	17.1	187	33.2	7.0	2.7	187	2.7	0.0	0.0	186	37.4	5.9	3.2	187	38.0*	3.2	1.6	187
	Heroin				Illicit Drugs[d]				Stimulants[e]				Sedatives[f]				Tranquilizers[g]			
Gay men																				
18–25	4.2	0.0	0.0	24	91.7	66.7	41.7	24	65.2	47.8	21.7*	23	29.2	8.3	4.2	24	29.2	16.7	0.0	24
26–34	1.9	0.0	0.0	106	81.5	42.6	25.9	108	30.8	8.4	3.7	107	18.7	2.8	0.0	107	25.2	6.5	1.9	107
35+	0.8	0.8	0.0	130	75.0	38.6*	26.5	132	33.1	4.6*	2.3*	130	24.8	5.4*	2.3	129	24.2	7.0	2.3	128
Total	1.5	0.4	0.0	260	79.2	42.8*	27.7	264	35.0	10.0	4.6*	260	22.7	4.6*	1.5	260	25.1	7.7	1.9	259
Lesbians																				
18–25	0.0	0.0	0.0	17	88.2	47.1	23.5	17	52.9	23.5	0.0	17	11.8	0.0	0.0	17	29.4	17.6	0.0	17
26–34	6.1	0.0	0.0	66	95.5*	40.9	24.2	66	54.5*	7.6	3.0	66	39.4*	1.5	1.5	66	38.5	6.2	1.5	65
35+	5.7*	0.0	0.0	105	81.3	24.3	16.8	107	39.4	1.9	0.0	104	33.3	0.0	0.0	105	34.3	6.7	1.9	105
Total	5.3*	0.0	0.0	188	86.8*	32.0	20.0	190	46.0*	5.9	1.1	187	33.5*	0.5	0.5	188	35.3*	7.5	1.6	187
	Analgesics[h]				Psychotherapeutics[i]				Alcohol				Cigarettes							
Gay men																				
18–25	25.0	10.8	3.2	24	75.0	54.2	29.2	24	100.0	100.0	83.3	24	79.2	62.5	58.3	24				
26–34	16.8	6.9	3.0*	107	37.0	14.8	5.6*	108	98.1	93.5	81.3*	107	56.1	33.6	26.2	107				
35+	17.9	5.8	0.9	123	41.7	9.1	3.0	132	98.5	82.4	70.2	131	73.8	40.8	35.4	131				
Total	18.1	7.1	2.2*	254	42.8	15.5	6.4*	264	98.5	88.5	76.0*	262	66.8	39.7	33.6	262				
Lesbians																				
18–25	17.6	5.9	0.0	17	58.8	35.3	0.0	17	100.0	94.1	76.5	17	82.4	70.6	52.9	17				
26–34	33.3*	7.6	0.0	66	54.5*	15.2	3.0	66	100.0	84.8	66.7	66	75.0*	50.0*	43.8*	64				
35+	21.0	4.8	0.0	105	51.4	10.3	1.9	107	99.1	84.0	63.2	106	78.1	42.9	38.1	105				
Total	25.0	5.9	0.0	188	53.2*	14.2	2.1	190	99.5	85.2	65.6	189	77.4*	47.8	41.4	186				

[a]Nonmedical use means without a doctor's prescription, in greater amounts or more often than prescribed, or for any other reason.
[b]Includes amyl or butyl nitrite and other inhalants.
[c]Includes LSD and other hallucinogens.
[d]Based on use of any illicit drug.
[e]Includes drugs such as dexedrine and benzedrine.
[f]Includes drugs such as quaaludes and seconal.
[g]Includes drugs such as Valium and Xanax.
[h]Includes drugs such as Darvon and Percodan.
[i]Based on nonmedical use of any psychotherapeutic drug.
*Differences in proportion between gay men and lesbians significant at .05 level.

largest cities located in a southern state. Participants were sampled by means of mailing lists from lesbian and gay organizations. A total of 455 respondents—190 (41.8%) lesbians and 265 (58.2%) gay men—completed the survey. This represents a 51.9% response rate overall. Mean age respective to each group was 36.4 and 37.4 years and mean years of education were 15.1 and 15.3 years. The majority of lesbians (92.3%) and gay men (94.1%) were White.

Within the context of research on lesbian and gay issues, the response rate, while low, is among the highest, and the coverage of illicit and licit drug use is among the most comprehensive reported to date.[13,16-19]

Measures

The drug use prevalence rates were calculated from responses to the question, "When was the most recent time you used [marijuana or hash], etc.?" The response categories were (1) never, (2) within the past week, (3) within the past month, (4) within the past year, (5) from 1 to 5 years ago, (6) from 5 to 10 years ago, and (7) 10 or more years ago. Past-year frequency of marijuana, alcohol, and cigarette use was measured by the question, "How often in the past year have you used [marijuana or hash], etc.?" Here the response categories were (1) never, (2) one to five times in the past year, (3) one to two times a month, (4) several times a month, (5) 1–2 days a week, and (6) daily or almost daily. These drugs were selected because they were the most common drugs used by both groups and had the largest variances. All drug questions were taken from the 1988 National Household Survey on Drug Abuse.[21]

Age was measured in number of years, and education was measured in

number of years of schooling. Personal income was categorized by (1) no personal income, (2) under $5000, (3) $5000 to $6999, (4) $7000 to $8999, (5) $9000 to $11 999, (6) $12 000 to $14 999, (7) $15 000 to $19 999, (8) $20 000 to $24 999, (9) $25 000 to $29 999, (10) $30 000 to $39 999, (11) $40 000 to $50 000, and (12) more than $50 000; household location was indicated by (1) rural, (2) town (under 10 000), (3) small city (10 000 to 100 000), (4) suburb, and (5) city (more than 100 000); and relationship status was measured by (0) single and (1) in relationship.

Statistical Procedures

To determine statistical differences in drug use between gay men and lesbians, a *z* test for differences in proportions for two samples was used. Ordinary least squares regression procedures were used to estimate the standardized regression coefficients.

Results

Prevalence of Illicit and Licit Drug Use

The prevalence of illicit and licit drug use for different age groups was similar for gay men and lesbians (Table 1). For inhalants and cigarettes, however, significant differences between gay men and lesbians occurred consistently: inhalant use was more prevalent among gay men and cigarette use was more prevalent among lesbians. Past-year and past-month use of any illicit drug, stimulants, and sedatives differed significantly between gay men and lesbians over age 35. The least prevalent drugs were crack and heroin.

Demographic Predictors of Marijuana, Alcohol, and Cigarettes

Linear multiple regression analysis indicates that for both groups, demographic predictors accounted for the most variance in cigarette use (Table 2). For lesbians, the frequency of marijuana use was significantly reduced with age and the frequency of cigarette use was significantly reduced with education; however, none of the demographic variables was a significant predictor of alcohol use. For gay men, education was important in reducing the frequency of both marijuana and cigarette use; living in urban areas increased the frequency of cigarette use; and income, household location, and relationship status significantly increased the frequency of alcohol use.

TABLE 2—Standardized Regression Coefficients (Linear Multiple Regression Analysis) for Demographic Predictors in Past-Year Self-Reports of Use of Marijuana, Alcohol, and Cigarettes

	Lesbians (n = 181)			Gay Men (n = 244)		
	Marijuana	Alcohol	Cigarettes	Marijuana	Alcohol	Cigarettes
Age	−.21**	−.11	−.07	−.10	−.03	.01
Education	−.10	.08	−.30**	−.17**	−.03	−.24**
Income	.06	.00	−.01	−.08	.15**	−.08
Household location	.04	−.06	−.09	−.03	.12*	.11*
Relationship status (single or in relationship)	.04	.03	−.04	.01	.12*	.00
R^2	.05*	.02	.12**	.07**	.05**	.09**

*Significant at .10 level.
**Significant at .05 level.

Discussion

Although a low response rate severely limits the interpretation of these data, they are justified by the absence of similar published data for both gays and lesbians living outside major metropolitan areas.

Marijuana and inhalants were the most popular illicit drugs reportedly used by gay men in this sample. For lesbians, cigarette use is particularly alarming. Rates for current smoking among women in national health studies are approximately 20 percentage points lower than those found here.[22]

Although frequency of drug use among lesbians and gay men declined with educational level, frequency of alcohol use increased among gay men who are affluent, reside in urban areas, and have close interpersonal relationships. These factors did not significantly affect drug use among lesbians.

Prevalence rates for this sample can be examined in light of the prevalence rates found in the 1990 National Household Survey on Drug Abuse.[23] Among adults aged 18 to 25, 16.5% of men and 9.1% of women have used marijuana in the past month, compared with 37.5% of gay men and 23.5% of lesbians. Among adults aged 26 to 34, 73.7% of men and 55.2% of women have used alcohol in the past month; comparative figures for gay men and lesbians are 81.3% and 66.7%, respectively. Finally, among adults aged 35 and over, 27.1% of men and 22.0% of women reported smoking cigarettes in the past month, compared with 35.4% for gay men and 38.1% for lesbians. Clearly, illicit and licit drug use in the homosexual community is a public health concern requiring immediate attention. □

Acknowledgments

This study was funded by the National Institute on Drug Abuse, grant R29-DA058321.

I would like to acknowledge Melanie Otis, Chris Miller, Natalie Saxe, and participating lesbian and gay organizations for their assistance in the project.

References

1. Saghir MT, Robins E, Walbran B, Gentry K. Homosexuality. IV: psychiatric disorders and disability in the female homosexual. *Am J Psychiatry*. 1970;127:147–154.
2. Diamond DL, Wilsnack SS. Alcohol abuse among lesbians: a descriptive study. *J Homosex*. 1978;4:123–142.
3. Swanson D, Loomis S, Cronin R, Smith J. Clinical features of the female homosexual patient: a comparison with heterosexual patients. *J Nerv Ment Dis*. 1972;155:119–124.
4. Anderson SC, Henderson DC. Working with lesbian alcoholics. *Soc Work*. 1985;30:518–525.
5. Lewis CW, Saghir MT, Robins E. Drinking patterns in homosexual and heterosexual women. *J Clin Psychiatry*. 1982;43:277–279.
6. Machover S, Puzzo F, Machover K, Plumeau F. Clinical and objective studies of personality variables in alcoholism: an objective study of homosexuality in alcoholism. *Q J Stud Alcohol*. 1959;20:528–542.
7. Gibbins R, Walter R. Three preliminary studies of a psychoanalytic theory of alcohol addiction. *Q J Stud Alcohol*. 1960;21:618–641.
8. McCord W, McCord J. *Origins of Alcoholism*. Stanford, Calif: Stanford University Press; 1960.
9. Rosenburg CM. Young alcoholics. *Br J Psychiatry*. 1969;181–188.
10. Roth LH, Rosenberg N, Levinson RB. Prison adjustment of alcoholic felons. *Q J Stud Alcohol*. 1971;32:382–392.
11. Fifield LH, Latham JD, Phillips C. *Alcohol-*

ism in the Gay Community: The Price of Alienation, Isolation, and Oppression. Los Angeles, Calif: Gay Community Services Center; 1977.
12. Saghir MT, Robins E, Walbran B, Gentry K. Homosexuality, III: psychiatric disorders and disability in the male homosexual. *Am J Psychiatry.* 1970;126:1079–1086.
13. Lohrenz LJ, Connelly JC, Coyne L, Spore KE. Alcohol problems in several midwestern homosexual communities. *J Stud Alcohol.* 1978;39:1959–1963.
14. Stall R, Wiley J. A comparison of drug and alcohol use habits of heterosexual and homosexual men. *Drug Alcohol Depend.* 1988;22:63–74.
15. Morales ES, Graves MA. *Substance Abuse: Patterns and Barriers to Treatment for Gay Men and Lesbians in San Francisco.* San Francisco, Calif: Dept of Public Health, City and County of San Francisco; 1983.
16. Martin J. Drinking patterns and drinking problems in a community sample of gay men. In: Seminara D, ed. *Alcohol and Immunomodulation and AIDS.* New York, NY: Alan R. Liss Inc; 1990:27–34.
17. McKirnan DJ, Peterson PL. Stress, expectancies, and vulnerability to substance use: a test of a model among homosexual men. *J Abnorm Psychol.* 1988;97:461–466.
18. McKirnan DJ, Peterson PL. Alcohol and drug use among homosexual men and women: epidemiology and population characteristics. *Addict Behav.* 1989;14:545–553.
19. McKirnan DJ, Peterson PL. Psychosocial and cultural factors in alcohol and drug abuse: an analysis of a homosexual community. *Addict Behav.* 1989;14:555–563.
20. Skinner WF, Otis M. Illicit/licit drug use among lesbians and gay people in a US southern state: epidemiological, comparative, and methodological findings from the Trilogy Project. *J Homosex.* In press.
21. *National Household Survey on Drug Abuse: Main Findings 1988.* Rockville, Md: National Institute on Drug Abuse; 1990.
22. *National Health Interview Survey, 1990.* Hyattsville, Md: National Center for Health Statistics; 1991.
23. *National Household Survey on Drug Abuse: Highlights 1990.* Rockville, Md: National Institute on Drug Abuse; 1991.

Breast Biopsy Rate and Positivity in Rhode Island

Margaret S. Richards, PhD, Judith Feldman, MD, MPH, Robert A. Smith, PhD, and Barbara A. DeBuono, MD, MPH

ABSTRACT

Breast biopsy is a common procedure for which few age-specific, population-based data are available. We identified all women who underwent breast biopsy in Rhode Island in 1989 by reviewing the specimen logs at each of 13 pathology laboratories. Among 425 000 women aged 15–97 years, 2685 underwent breast biopsy, for an overall rate of 6/1000. The rate increased with age, peaking among 45- to 54-year-olds and then declining in women over 75. Among those biopsied, 726 were diagnosed with breast cancer, for an overall biopsy positivity of 27%. In contrast to rate, positivity increased steadily with age. These results are within the range of estimates produced by smaller group studies. (*Am J Public Health.* 1994;84:1310–1312)

Introduction

In November 1987, the Rhode Island Department of Health, in collaboration with the Centers for Disease Control, initiated a Breast Cancer Screening Program. A review of data from the quality assurance component of this program raised questions about the extent of presurgical workup performed before biopsy. Although a number of publications have addressed this issue,[1–3] the majority have come from centers with special interest and expertise in mammography.[4] To date, we are unaware of any population-based studies of the association between prebiopsy evaluation and breast biopsy outcome.

The Rhode Island Department of Health has undertaken a two-phase, population-based study of breast biopsy among Rhode Island women. Phase I, which is complete, involved identifying all female residents who underwent open breast biopsy in the state in the calendar year 1989. The purpose of this phase was to generate a sampling frame for Phase II, which involves abstracting medical record data for a sample of women undergoing biopsy to identify those lesions and those diagnostic pathways that consistently lead to a negative (potentially avoidable) biopsy. This paper summarizes the descriptive data obtained in Phase I on biopsy rates and positivity.

Methods

Definitions

An open breast biopsy was defined as the removal of breast tissue to provide a definitive histologic diagnosis. This definition excludes percutaneous or cytologic procedures such as fine needle aspiration, treatment of established breast cancer such as mastectomy, and other procedures involving the breast, such as breast

At the time of the study, Margaret S. Richards was with the Office of Disease Control, Rhode Island Department of Health, Providence, RI, and the Epidemiology Program Office, Centers for Disease Control and Prevention, Atlanta, Ga; she is now with Abbott Laboratories, in Abbott Park, IL. Judith Feldman is with the Division of Preventive Health Services, Rhode Island Department of Health. Robert A. Smith was with the Division of Cancer Prevention and Control, Centers for Disease Control and Prevention; he is now with the Medical Affairs Department, American Cancer Society, Atlanta, Ga. Barbara A. DeBuono is with the Rhode Island Department of Health.

Requests for reprints should be sent to Judith Feldman, MD, MPH, Division of Preventive Health Services, 3 Capitol Hill, Rm 403, Providence, RI 02908-5097.

This paper was accepted November 12, 1993.

reduction or reconstruction. A positive biopsy is defined as a biopsy resulting in the diagnosis of breast cancer.

Enrollment

Women were included in the study if they had breast tissue examined at one of 13 hospital-based pathology laboratories processing breast specimens in the state of Rhode Island. The 13 laboratories process all outpatient and inpatient specimens. Surgical specimen logbooks were reviewed for any entry labeled "breast," and a specimen number was obtained. The corresponding pathology report was read. Breast specimens obtained from males or from procedures other than open biopsy were excluded.

Data Collection

The information obtained for each woman included name, birth date or age, biopsy date, biopsy outcome (positive or negative), histologic outcome (e.g., infiltrating ductal carcinoma), and hospital code (1–13). Rates of biopsy were calculated by using 1989 Census Bureau population estimates as a denominator.

Results

Biopsy Rates

Among 425 000 women age 15 to 97 years, 2685 underwent open breast biopsy in 1989, for an overall biopsy rate of 6/1000. The rate of biopsy increased with age, peaking among 45- to 54-year-olds and then declining in women over the age of 75 (Figure 1). Seventy percent of the women biopsied were diagnosed as having a benign breast condition (fibrocystic disease accounted for 72% of these diagnoses). Four percent of the women had noninvasive cancer, whereas 23% had invasive cancers; in both categories, ductal carcinoma was the dominant histologic type. For 3% of the biopsies, the lesions involved malignant tissue that was not of breast origin.

Biopsy Positivity

Among 2685 women biopsied, 726 were diagnosed with breast cancer, for an overall biopsy positivity of 27%. Positivity increased steadily with age, so that 75% of the biopsies performed on women age 85 and over revealed cancer (Figure 2). This steady increase with age in biopsy *positivity* is in contrast to the increase followed by a decline in biopsy *rate*.

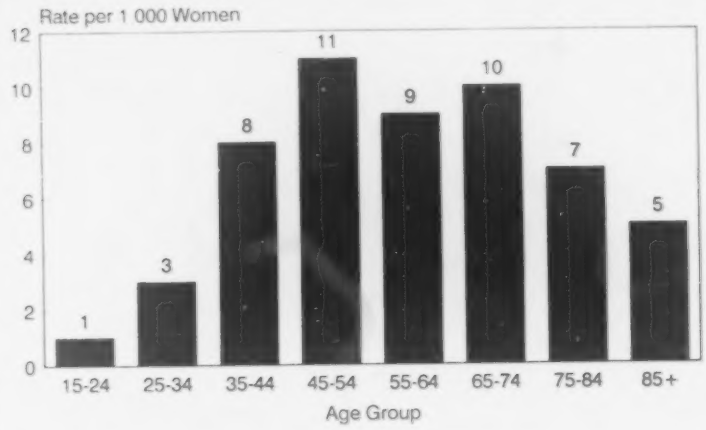


FIGURE 1—Breast biopsy rates by age, Rhode Island, 1989.

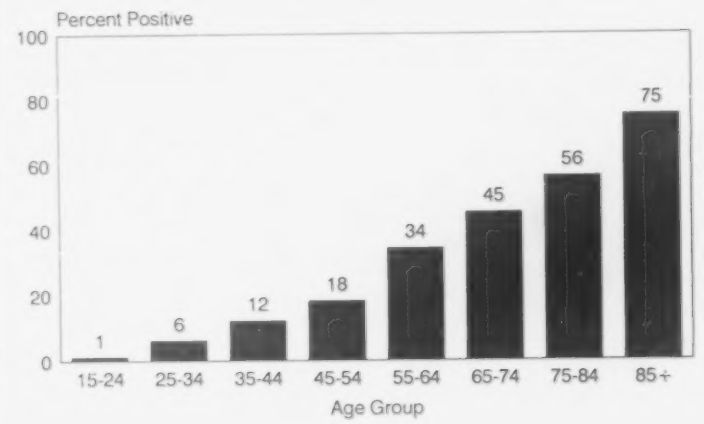


FIGURE 2—Breast biopsy positivity by age, Rhode Island, 1989.

Discussion

Efforts to achieve a consensus on the ideal breast biopsy positivity ratio have been hindered, in part, by a lack of population-based data on this procedure. However, some patterns emerge from these data that are in agreement with other studies. The overall biopsy positivity of 27% in our clinically heterogeneous group of women appears to be consistent with previous estimates. In other series,[3,5] positive biopsy percentages ranged from 9% to 65%, with a 15% to 30% positive biopsy rate among asymptomatic women[16] and a 20% to 40% rate among symptomatic women.[17–19] Our finding that fibracystic disease accounted for 72% of all the benign diagnoses is similar to that of a previous study.[20]

It is noteworthy that the older women (who are known to be at higher risk for breast cancer) have lower rates of biopsy but a higher rate of positivity in our study. The lower rate of biopsy may reflect a reluctance on the part of the patient or the surgeon to schedule a traumatic surgery unless the clinical suspicion of cancer is high (i.e., the woman is symptomatic). Cancer is more easily diagnosed in

the older, less dense breast, however—a potential factor in the increased positivity.

In Phase I of our study we essentially established a breast biopsy "registry" for the year 1989. A registry such as this one is population based if it includes Rhode Island residents biopsied out of state but excludes out-of-state residents biopsied in Rhode Island. Only a few of Rhode Island's hospitals include patient address on the pathology report, so we were unable to include residents diagnosed out of state and exclude nonresidents diagnosed in state.

We used the Rhode Island Cancer Registry to provide a proxy measure of breast biopsy referral bias in and out of state. In 1989, 926 cases of incident breast cancer were reported to the registry. Of these 926 cases, 29 (3%) cases were Rhode Island residents diagnosed in either Massachusetts or Connecticut and reported to Rhode Island through a reciprocal agreement. Fifty-five (6%) of the 926 cases were Massachusetts or Connecticut residents diagnosed in this state.

We theorize that most residents undergo biopsy in their home state, having been referred to a local surgeon by their primary care physician. If the biopsy is positive, a woman may choose to go out of state (e.g., Boston) for what she perceives to be specialist treatment. Because a positive biopsy is only one of several pathways to a cancer registry report, we assume that referral bias in our biopsy registry is minimal.

With these data, we are able to measure the parameters of biopsy rate and biopsy positivity and how these parameters vary across subgroups of age. After the completion of Phase II, we will be able to estimate how these parameters are mitigated by the family history of the patient and the clinical history of the patient, including symptomatology, histology, and prebiopsy diagnostic evaluation. □

References

1. *The Breast Cancer Digest: A Guide to Medical Care, Emotional Support, Educational Programs, and Resources.* 2nd ed. Bethesda, Md: National Cancer Institute; April 1984. NIH publication PHS 84-1691.
2. Danforth DN, Lichter AS, Lippman ME. The diagnosis of breast cancer. In: Lippman ME, Lichter AS, Danforth DN, eds. *Diagnosis and Management of Breast Cancer.* Philadelphia, Pa: WB Saunders Co; 1988:50–94.
3. Mushlin AI. Diagnostic tests in breast cancer: clinical strategies based on diagnostic probabilities. *Ann Intern Med.* 1985;103: 79–85.
4. Morrow M. Management of nonpalpable breast lesions. *Principles and Practice of Oncology.* 1990;4(1):1–11.
5. Choucair R, Holcomb M, Matthews R, Hughes T. Biopsy of nonpalpable breast lesions. *Am J Surg.* 1988;156:453–456.
6. Tabar L, Gad A. Screening for breast cancer: the Swedish trial. *Radiology.* 1981; 138:219–222.
7. Schwartz G, Feig S, Patchefsky A. Significance and staging of nonpalpable carcinomas of the breast. *Surg Gynecol Obstet.* 1988;166:6–10.
8. Ciatto S, Cataliotti L, Distante V. Nonpalpable lesions detected with mammography: review of 512 consecutive cases. *Radiology.* 1987;165:99–102.
9. Cherry U, Kirkpatrick A, Anderson T, et al. Localization and excision of occult breast lesions. *Br J Surg.* 1983;70:607–610.
10. Bigelow R, Smith R, Goodman P, Wilson G. Needle localization of nonpalpable breast masses. *Arch Surg.* 1988;120:565–569.
11. Skinner M, Swain M, Simmons R. Nonpalpable breast lesions at biopsy. *Ann Surg.* 1988;203–208.
12. Dershaw D. Nonpalpable, needle-localized mammographic abnormalities: pathologic correlation in 219 patients. *Cancer Invest.* 1986;1–4.
13. Gisvold J, Martin J. Pre-biopsy localization of nonpalpable breast lesions. *Am J Roentgenol.* 1984;143:477–481.
14. Landercasper J, Gundersen S, Gundersen A, et al. Needle localization and biopsy of nonpalpable lesions of the breast. *Surg Gynecol Obstet.* 1987;164:399–403.
15. Rubin E, Visscher D, Alexander R, et al. Proliferative disease and atypia in biopsies performed for nonpalpable lesions detected mammographically. *Cancer.* 1988;61: 2077–2082.
16. Meyer J, Kopans D, Stomper P, Lindfors K. Occult breast abnormalities: percutaneous pre-operative needle localization. *Radiology.* 1984;150:335–337.
17. Winchester DP, Sener S, Immerman S, et al. A systematic approach to the evaluation and management of breast masses. *Cancer.* 1983;51:2535–2539.
18. Kagali VA. The role and limitations of frozen section diagnosis of a palpable mass in the breast. *Surg Gynecol Obstet.* 1983;156: 168–170.
19. Coates MR, Pilch YH, Benfield JR. Changing concepts in establishing the diagnosis of breast masses. *Am J Surg.* 1977;134:77–81.
20. Cole P, Elwood JM, Kaplan SD. Incidence rates and risk factors of benign breast neoplasms. *Am J Epidemiol.* 1978;108:112–120.

Projected Numbers of Cancers Diagnosed in the US Elderly Population, 1990 through 2030

Anthony P. Polednak, PhD

ABSTRACT

As based on Bureau of the Census population projections and age-specific cancer incidence rates for 1985 to 1989 from the National Cancer Institute's Surveillance, Epidemiology and End Results program, the number of incident cancers diagnosed annually in the United States among persons aged 65 and over is projected to reach 1.5 million by the year 2030, or 2.4 times the number estimated for 1990. These projections, which may be conservative because birth cohort patterns (based on Connecticut rates) suggest possible future increases in incidence rates for all cancer sites combined, have implications for planning expanded primary prevention efforts, such as smoking cessation (especially for women) and dietary modification programs, and for projecting health care needs and costs. (*Am J Public Health.* 1994;84:1313–1316)

Introduction

Anticipated increases in the number of elderly persons in the US population in the coming decades, especially those aged 80 years and over, will result in increases in the numbers of incident cancers diagnosed.[1,2] This is especially true for cancers that tend to occur mainly in the elderly (e.g., those of the colon, rectum, pancreas, stomach, lung, bladder, and prostate).[1] This report uses available data on population projections and cancer incidence rates to estimate the total numbers of incident cancer cases that might be diagnosed in selected years between 1990 and 2030, and it discusses some implications of the projected increases.

Methods

Population projections were available from the US Bureau of the Census by age and sex; the middle series assumes constant fertility rates (with minor exceptions), slow increases in life expectancy, and constant net immigration.[3] Cancer incidence rates were obtained from the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) program, which includes nine population-based cancer registries (Connecticut, Hawaii, Iowa, New Mexico, Utah, Detroit, Atlanta, San Francisco–Oakland, and Seattle–Puget Sound) covering about 9.5% of the US population; these data are regarded as "reasonably representative" of the US population.[4] Rates were for all invasive cancers (along with in situ bladder tumors); leukemias and lymphomas were included.[4] SEER age- and sex-specific average annual incidence rates for all cancers diagnosed between 1985 and 1989 at ages 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85+ years[4] were multiplied by population estimates for the same age groups for each sex[3] for every fifth calendar year between 1990 and 2030. Projected numbers refer to cancers and not to numbers of patients with cancer.

Results

The numbers of incident cancers (all sites combined) are projected to increase more rapidly for men than for women in the next few decades (Table 1 and Figure 1). This is owing to greater predicted increases in the number of elderly men versus women in the US population, reflecting faster increases in men's life expectancy.[3] While the number of cancers diagnosed annually among persons aged 80 and over was slightly greater for women than for men in 1990, a crossover may occur by about 2010 (Table 1 and Figure 1).

After 2010, with the aging of the baby boom generation, the number of elderly persons will increase more rapidly than in previous years, resulting in larger increases in projected numbers of incident cancers. Percentage increases in the projected numbers of incident cancers per year, estimated every 5 years, were smaller from 1990 to 2010 than from 2010 to 2030.

Discussion

This study has several limitations. First, cancer rates for a broad age group were used for the oldest age category (85+ years) because data on cancer rates within this age group were lacking; this age group included 3.1 million persons (1.2% of the population) in 1990 and will comprise an estimated 8.4 million persons (2.4% of the population) in 2030.[3] Projected numbers of cancers in those aged 85 and over could be inaccurate owing both to rapid increases in incidence rates for certain cancers within this group and to greater increases in population size among the oldest ages. Second, if improvements in medical care (including cancer

The author is with the Connecticut Department of Public Health and Addiction Services, Hartford, Conn.

Requests for reprints should be sent to Anthony P. Polednak, PhD, Connecticut Tumor Registry, Connecticut Department of Public Health and Addiction Services, 150 Washington St, Hartford, CT 06106.

This paper was accepted February 15, 1994.

TABLE 1—Projected Numbers of Incident Cancers for US Men and Women Aged 65 and Over and Aged 80 and Over

	Age 65 and Over				Age 80 and Over			
	Men		Women		Men		Women	
Year	No.	% Change[a]	No.	% Change	No.	% Change	No.	% Change
1990	344 213	...	305 817	...	82 611	...	92 672	...
1995	380 177	10	328 595	7	96 693	17	106 359	15
2000	407 118	7	341 510	4	112 277	16	119 255	12
2005	433 989	7	355 484	4	131 308	17	133 485	12
2010	477 294	10	377 282	6	145 376	11	142 234	7
2015	555 007	16	422 941	12	153 794	6	145 278	2
2020	660 793	19	484 196	14	168 792	10	152 189	5
2025	787 573	19	558 096	15	201 920	20	172 322	13
2030	905 593	15	626 870	12	258 876	28	210 127	22

Note. Only invasive cancers are included, except for in situ bladder tumors.
[a]Percentage change in number of incident cancers diagnosed annually (e.g., the number of cases diagnosed in 2005 minus the number in 2000, divided by the number in 2000).

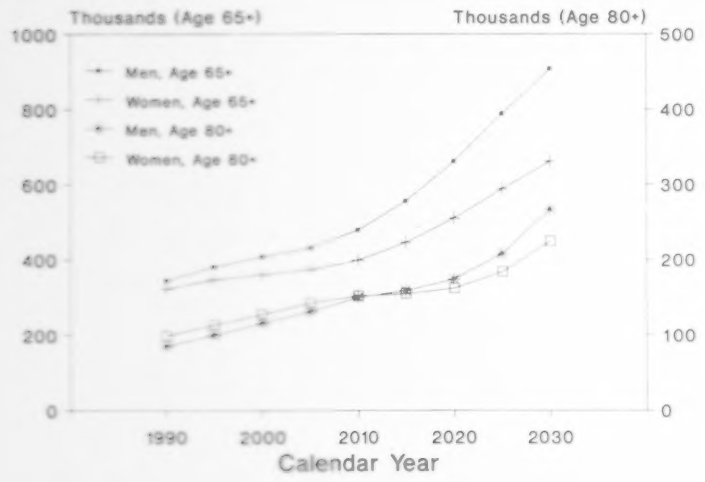


FIGURE 1—Projected numbers of incident cancers in the US population aged 65 and over and aged 80 and over, in selected calendar years.

treatments) result in faster increases in life expectancy than those assumed by the middle series of population projections, increases in the elderly population would be greater (as in the "high life expectancy" series)[3] than those assumed in this paper, and projected numbers of incident cancers would be greater than those in Table 1. Third, use of SEER rates probably resulted in a slight underestimation of incidence rates[5] and, hence, of projected numbers of cancers. The effect of projected increases in the proportion of minority members (e.g., Blacks, Hispanics, and Asians–Pacific Islanders) in the US population in the coming decades[1] was not considered, but it may be small. This is because cancer incidence rates have been higher for Blacks than for Whites[6] but lower for Hispanics (the fastest growing minority population), despite documented secular increases in cancer incidence rates in some Hispanic populations.[7]

In a 1984 report,[2] about 1.5 million cases of cancer for all ages combined were predicted for the year 2030; numbers for the elderly alone were not given.[2] The present analyses predict 1.5 million incident cases in 2030 for age 65 and over alone (Table 1); the projected number for all ages would be 2.1 million (data not shown).

Secular trends in cancer incidence and mortality rates often reflect birth cohort patterns because of cohort-related changes in exposure to carcinogens.[8–10] Birth cohort patterns were examined for Connecticut residents because cancer incidence rates in that state are available for years prior to the establishment of the SEER program in 1973 and rates for the elderly may closely approximate those for the entire United States.[5] Cohort patterns suggest that age-specific cancer rates (in 5-year intervals from ages 50 to 54 to ages 80 to 84) for all cancers combined in elderly men (Figure 2) and women (Figure 3) are unlikely to decrease—and could increase—in the next few decades. Thus, the projected numbers of cancers in the elderly (Table 1), based on the assumption of constant incidence rates over time, are probably conservative (i.e., underestimated), at least for the next few decades.

The latter conclusion is supported by the fact that few cancer sites have shown marked declines over time or across birth cohorts. Stomach cancer incidence rates have declined across birth cohorts (not shown), probably reflecting changes in diet[11] and exposure to specific bacteria,[12] but absolute rates were low by 1985 to 1989 (with no further decline in age-standardized rates in 1990–91), and these cancers have only a small effect on total cancer rates. For lung cancer, Connecticut incidence data (not shown) and US mortality data[13] indicate that rates, strongly influenced by sex- and cohort-specific smoking habits,[9,13–15] peaked for men born around 1925 to 1930 and for women born 10 years later and are declining in later cohorts (represented, as yet, only by nonelderly age groups). Any effect of declines in lung cancer rates (as post-1930 cohorts of men reach progressively older ages) on projected numbers of total cancers in elderly men, however, could be offset by future increases in incidence rates for prostate cancer (expected from birth cohort patterns). Even if some of the increasing numbers of prostate cancer cases detected by screening are "indolent" or clinically silent, these cases involve diagnostic procedures and often receive treatment.

The projected numbers of total incident cancers among women (Table 1) are probably conservative because women

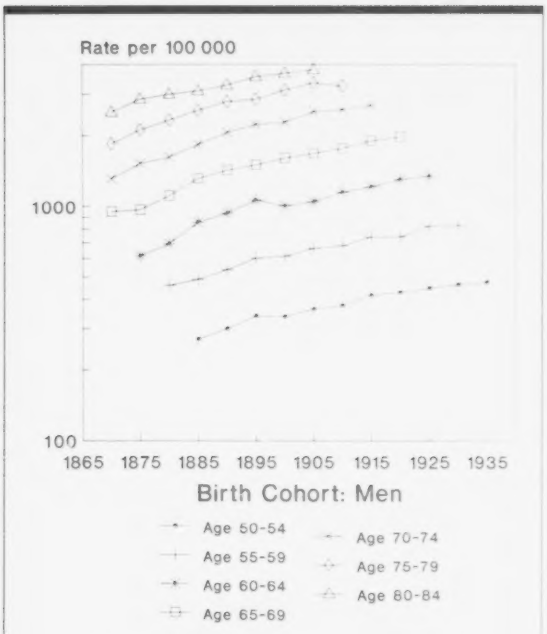


FIGURE 2—Age-specific average annual incidence rates per 100 000 men for cancers of all sites, by birth cohort (central year of birth) in Connecticut residents.

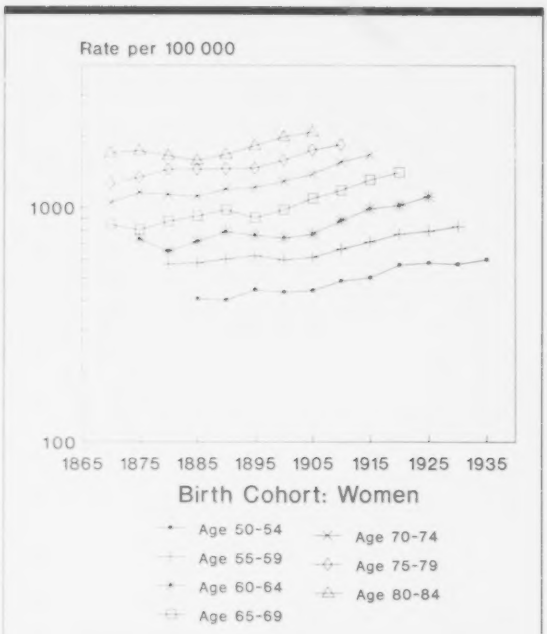


FIGURE 3—Age-specific average annual incidence rates per 100 000 women for cancers of all sites, by birth cohort (central year of birth) in Connecticut residents.

born around 1945 and later, for whom lung cancer incidence rates have started to decline, will not reach age 65 until 2010 (and will not reach age 80 until 2025). Although the use of breast cancer incidence rates for 1985 to 1989 in this study could be problematic because age-standardized incidence rates for invasive breast cancer increased more between that period and the 5-year interval preceding it than they did in previous periods (probably reflecting trends in screening), incidence rates for invasive breast cancer are not declining but are instead returning to a slower rate of increase like that evident in Connecticut since about 1940[8,16] and across birth cohorts until at least the 1955 cohort.[10]

Future studies should include projections of incidence rates by cancer site based on age-period-cohort modeling. The crude projections in Table 1 (and Figure 1), however, underscore the importance of primary prevention to reduce the impact of an aging population on the cancer burden. The birth cohort pattern for lung cancer incidence and mortality rates in women, along with evidence for higher risk in female than in male smokers,[17] supports the expansion of smoking cessation and prevention programs targeted to women.[18] A secular trend toward reduced fat intake may be occurring,[19,20] but intensifying efforts to reduce fat consumption and increase fruit and vegetable intake, as embodied in national objectives for the year 2000,[21,22] should eventually affect incidence rates for various cancers (as well as cardiovascular diseases) favorably, given the effects of antioxidant nutrients in certain foods (and nutritional supplements).[23] Along with chemopreventive effects of nutritional changes, progress toward national objectives for reducing alcohol and tobacco use[21,22] would decrease the risks for primary cancers of the aerodigestive tract and second primaries in patients with these cancers.[24]

The aging of the population was only a minor factor in the growth of health care costs from 1950 to 1987.[25] However, even if health expenditures increased only as fast as the general economy by 1996, the aging of the population (by itself) could cause health spending to reach 16% of the gross national product in 2030 (a 33% increase over the 12% in 1990).[26] There is increasing recognition of the need for studies of cancer treatment effectiveness in elderly cancer patients, who are often excluded from clinical trials.[1,27] While objective data on the cost-effectiveness of cancer treatment options in an aging population[28,29] are also being sought, equity (by age) in the receipt of effective cancer therapy should be the paramount goal, and patient preferences[30,31] cannot be ignored. □

Acknowledgments

This work was supported by contract N01–CN–05226 between the National Cancer Institute and the Connecticut Department of Health Services.

References

1. Monfardini S, Yancik R. Cancer in the elderly: meeting the challenge of an aging population. *J Natl Cancer Inst.* 1993;85:532–538.
2. Janerich DT. Forecasting cancer trends to optimize control strategies. *J Natl Cancer Inst.* 1984;72:1317–1321.
3. Population projections of the United States by age, sex, race and Hispanic origin: 1992 to 2050. *Current Population Reports,* Series P-25, no. 1092. Washington, DC: US Bureau of the Census; 1992.
4. Miller BA, Gloeckler Ries LA, Hankey BF, et al., eds. *Cancer Statistics Review 1973–1989.* Bethesda, Md: National Can-

cer Institute; 1992. NIH publication 92-2789.
5. McBean MA, Babish JD, Warren JL. Determination of lung cancer incidence in the elderly using Medicare claims data. *Am J Epidemiol.* 1993;137:226–234.
6. Baquet CR, Horm JW, Gibbs T, Greenwald P. Socioeconomic factors and cancer incidence among blacks and whites. *J Natl Cancer Inst.* 1991;83:551–557.
7. Savitz DK. Changes in Spanish surname cancer rates relative to other whites, Denver area, 1969–71 to 1979–81. *Am J Public Health.* 1986;76:1210–1215.
8. Roush GC, Holford TR, Schymura MJ, White C. *Cancer Risk and Incidence Trends: The Connecticut Perspective.* New York, NY: Hemisphere Publishing Corp; 1987.
9. Boyle P, Robertson C. Statistical modeling of lung cancer and laryngeal cancer incidence in Scotland, 1960–79. *Am J Epidemiol.* 1987;125:731–744.
10. Holford TR, Roush GC, McKay LA. Trends in female breast cancer in Connecticut and the United States. *J Clin Epidemiol.* 1991;44:29–39.
11. Howson CP, Hiyama T, Wynder EL. The decline in gastric cancer: epidemiology of an unplanned triumph. *Epidemiol Rev.* 1986;8:1–27.
12. Forman D. Geographic association of Helicobacter pylori antibody prevalence and gastric cancer mortality in rural China. *Int J Cancer.* 1990;46:608–611.
13. Burbank F. US lung cancer death rates begin to rise proportionately more rapidly for females than for males: a dose-response effect? *J Chron Dis.* 1972;25:473–479.
14. Shopland DR, Eyre HJ, Pechacek TF. Smoking-attributable cancer mortality in 1991: is lung cancer now the leading cause of death among smokers in the United States? *J Natl Cancer Inst.* 1991;83:1142–1148.
15. Devesa SS, Blot WJ, Fraumeni JF, Jr. Declining lung cancer rates among young men and women in the United States: a cohort analysis. *J Natl Cancer Inst.* 1989;81:1568–1572.
16. Feuer EJ, Wun LM. How much of the recent rise in breast cancer incidence can be explained by increases in mammography utilization? *Am J Epidemiol.* 1992;136:1423–1436.
17. Risch HA, Howe GR, Jain M, et al. Are female smokers at higher risk for lung cancer than male smokers? A case-control analysis by histologic type. *Am J Epidemiol.* 1993;138:281–293.
18. Ernster VL. Women and smoking. *Am J Public Health.* 1993;83:1202–1204.
19. Stephen AM, Wald NJ. Trends in individual consumption of dietary fat in the United States, 1920–1984. *Am J Clin Nutr.* 1990;52:457–469.
20. Nicklas TA, Webber LS, Srinivasan SR, Berenson GS. Secular trends in dietary intakes and cardiovascular risk factors of 10-y-old children: the Bogalusa Heart Study (1973–1988). *Am J Clin Nutr.* 1993;57:930–937.
21. Greenwald P, Sonkik EJ. *Cancer Control Objectives for the Nation: 1985–2000.* Bethesda, Md: National Cancer Institute; 1986;2. NIH publication 86–2880.
22. Public Health Service. *Healthy People 2000: National Health and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50212.
23. Blot WJ, Li JY, Taylor PR, et al. Nutrition intervention trials in Linxian, China: supplementation with specific vitamin/mineral combinations, cancer incidence and disease-specific mortality in the general population. *J Natl Cancer Inst.* 1992;85:1483–1492.
24. Day GL, Blot WJ, Shore RE, et al. Second cancers following oral and pharyngeal cancers: role of tobacco and alcohol. *J Natl Cancer Inst.* 1994;86:131–137.
25. Aaron H, Schwartz WB. Rationing health care: the choice before us. *Science.* 1990; 247:418–422.
26. Waldo DR, Sonnefeld ST, Lemieux JA, McKusick DR. Health spending through 2030: three scenarios. *Health Affairs.* 1991; 10:231–242.
27. Harris JE. The treatment of cancer in an aging population. *JAMA.* 1992;268:96–97.
28. Parker RG. Varying charges for comparably effective cancer treatments. *Am J Clin Oncol.* 1992;15:281–288.
29. Smith TJ, Hillner BE, Desch CE. Efficacy and cost-effectiveness of cancer treatment: rational allocation of resources based on decision analysis. *J Natl Cancer Inst.* 1993; 85:1460–1474.
30. Ganz PA. Treatment options for breast cancer—beyond survival. *N Engl J Med.* 1992;326:1147–1149.
31. Newcombe PA, Carbone PP. Cancer treatment and age: patient perspectives. *J Natl Cancer Inst.* 1993;85:1580–1584.

Subsequent Hip Fracture among Older Adults

ABSTRACT

Subsequent hip fracture among the 368 Longitudinal Study on Aging respondents who fractured their hips from 1984 through 1991 was prospectively examined. Case-by-case review of the billing records indicated that 27 subsequent hip fractures occurred, for a rate of 1 every 33.8 person-years. Multivariable proportional hazard regression revealed that increased risks of subsequent hip fracture were associated with poor perceived health status and dizziness. (*Am J Public Health.* 1994; 84:1316–1318)

Fredric D. Wolinsky, PhD, and John F. Fitzgerald, MD, MBA

Introduction

Hip fractures occur frequently among older adults. Studies suggest that the lifetime risk of hip fracture is 6.0% for men and 17.5% for women.[1] Virtually all hip fractures require hospitalization.[2] Sequelae can include substantially reduced functional status or death.[3–5] Accordingly, a key target in federal health care policy is the reduction of the incidence of hip fracture in the elderly.[6]

Although research on the risk of hip fracture has increased,[7–11] little is known about subsequent hip fracture.[12] Indeed, there are no national studies of subsequent hip fracture. Accordingly, the purpose of this report is to prospectively assess the risk of subsequent hip fracture among the 368 Longitudinal Study on Aging respondents who were hospitalized for hip fracture from 1984 through 1991.[11]

Methods

Prior to our earlier report,[11] no comprehensive conceptual model of the

The authors are with the Department of Medicine, Indiana University School of Medicine, and the Regenstrief Institute for Health Care, Indianapolis, Ind. John F. Fitzgerald is also with the Richard Roudebush Veterans Administration Medical Center, Indianapolis.

Requests for reprints should be sent to Fredric D. Wolinsky, PhD, Department of Medicine, Indiana University School of Medicine, 1001 W Tenth St, RG-5, Indianapolis, IN 42020-2859.

This paper was accepted March 30, 1994.

Note. The opinions expressed here are the authors' and do not necessarily reflect those of the funding agency, academic institutions, or government institutions involved.

risk of hip fracture existed. We classified the known and plausible risk factors associated with a first hip fracture into (1) ecological factors, (2) demographic factors, (3) social factors, (4) general health status, (5) symptoms and diseases conducive to hip fracture, (6) falling history and body mass, and (7) previous hip fracture. This sequence corresponds to a natural progression from the most distal to the most proximal risks, and it is consistent with the consensus that the most immediate risk factors are falling history, body mass, and prior hip fracture.[3,12,13] Although the risk of subsequent hip fracture may be dictated by a different process, initial reliance on the model developed for the first hip fracture is appropriate.

The data are taken from the Longitudinal Study on Aging, which is the follow-up to the Supplement on Aging appended to the 1984 National Health Interview Survey. The Longitudinal Study on Aging consists of the 7527 Supplement on Aging individuals aged 70 years or more in 1984 who were selected for follow-up interviews. Abstracts of the Medicare Automated Data Retrieval System, Part A (i.e., hospitalization), and National Death Index files are also available. We linked these files, identified the hospitalization episodes that contained the *International Classification of Diseases*, 9th edition, Clinical Modification (ICD-9-CM) codes for hip fracture (i.e., 820.0 through 820.9) in any of the five discharge diagnosis fields, and selected the first such episode from 1984 through 1991 for each Longitudinal Study on Aging respondent. These 368 individuals constitute our sample. Where multiple hip fracture billing records existed, we also selected all ensuing billings. These suggest subsequent hip fracture. Case-by-case review was then conducted to verify that these were actual subsequent hip fractures and not duplicate billings, same-day transfers to another hospital, or complications stemming from the first hip fracture. We used the unweighted Longitudinal Study on Aging data without adjustments for design effects[14–17] and the same variables we used in our analysis of first hip fractures.[11] These are defined and their means or percentages are provided in Table 1. To reduce bias due to overfitting or multicollinearity, all bivariate relationships between subsequent hip fracture and each independent variable were examined with a screening criterion of $P \leq 10$.[18] Cox proportional hazard regression was used. The date of the Longitudinal Study on Aging respondent's first hip fracture marked the beginning of the period of observation, and the date of subsequent hip fracture (censored at the end of the observation period—December 31, 1991) marked the end. Decedents were censored at the National Death Index date of their death. All analyses were performed with SPSS for Windows.[19]

TABLE 1—Independent Variables on the Risk of Subsequent Hip Fracture among the 368 Longitudinal Study on Aging Respondents Who Fractured Their Hips at Least Once, 1984 through 1991

Independent Variable	Mean or Percentage of Sample	Crude Rate Ratio with Subsequent Hip Fracture	95% CI
Ecological factors			
Rural residence[a]	37.2%	1.04	0.48, 2.24
Southern residence[a]	34.0%	1.74	0.81, 3.73
Demographic characteristics			
Age[b]	79.7	0.99	0.93, 1.06
Female gender[a]	78.0%	1.16	0.44, 3.07
Black race[a]	3.0%	2.82	0.66, 11.97
Social factors			
Lives alone[a]	47.0%	1.35	0.63, 2.88
Education[c]	1.8	0.72	0.42, 1.24
Kin social supports[d]	1.6	1.28	0.65, 2.54
Nonkin social supports[e]	2.2	1.12	0.84, 1.48
General health status			
Poor perceived health status[f]	35.3%	2.21	1.03, 4.72
Hospitalized in the year before baseline[a]	28.3%	1.39	0.63, 3.03
Symptoms and diseases conducive to hip fracture			
Visual problems[a]	13.0%	1.74	0.06, 4.61
Hearing problems[a]	9.8%	1.99	0.68, 5.76
Dizziness[a]	32.6%	2.58	1.20, 5.51
Difficulty walking ¼ mile[a]	39.4%	0.85	0.38, 1.90
Difficulty walking up 10 steps[a]	31.5%	0.78	0.33, 1.84
Difficulty standing for 2 hours[a]	47.0%	0.52	0.23, 1.20
Difficulty sitting for 2 hours[a]	11.1%	1.01	0.30, 3.35
Difficulty stooping, crouching, or kneeling[a]	45.4%	0.70	0.32, 1.54
Difficulty reaching over the head[a]	22.3%	0.68	0.24, 1.97
Difficulty shaking hands[a]	3.3%	1.72	0.23, 12.79
Difficulty using fingers to grasp[a]	12.2%	0.87	0.26, 2.87
Difficulty carrying 25 pounds[a]	49.7%	1.43	0.67, 3.06
Ever had osteoporosis[a,g]	4.6%	1.82	0.43, 7.71
Ever had a stroke[a,g]	6.3%	0.56	0.08, 4.10
Ever had diabetes[a,g]	6.5%	0.65	0.09, 4.77
Ever had arthritis[a,g]	57.1%	0.63	0.30, 1.34
Falling history and body mass			
Falling status in the year before baseline[a]	33.7%	0.55	0.22, 1.37
Body mass[h]	23.2	0.97	0.87, 1.07

[a]Coded yes = 1, no = 0.
[b]Coded in years.
[c]Coded grade school = 1, high school = 2, college = 3.
[d]A two-item scale coded as both supports = 2, one support = 1, no supports = 0.
[e]A five-item scale coded as five supports = 5, four supports = 4, . . . one support = 1, no supports = 0.
[f]Coded fair or poor = 1, excellent, very good, or good = 0.
[g]Based on self-reports, which are specific but not sensitive.
[h]Weight (kg)/height^2 (m).

Results

Of the 368 Longitudinal Study on Aging respondents hospitalized for hip fracture between January 1, 1984, and December 31, 1991, 51 had multiple billing records suggestive of subsequent hip fracture. Case-by-case review, however, revealed that 3 were duplicate

billings, 14 were same-day transfers, and 7 were rehospitalizations for medical or surgical complications from the first hip fracture. Thus, only 27 (7.3%) of these 368 respondents had subsequent hip fractures during the study period. This is the equivalent of one subsequent hip fracture every 33.8 person-years.

Table 1 contains the crude rate ratios (RRs). Only two variables met the screening criterion: poor (fair or poor vs good, very good, or excellent) perceived health (RR = 2.21; 95% confidence interval [CI] = 1.03, 4.72), and reported problems with dizziness (RR = 2.58; 95% CI = 1.20, 5.51). When advanced to the multivariable proportional hazards regression, neither variable had a statistically significant independent effect (poor perceived health: adjusted rate ratio [ARR] = 1.76; 95% CI = 0.80, 3.90; problems with dizziness: ARR = 2.18; 95% CI = 0.98, 4.83), although the equation was statistically significant, suggesting some collinearity between these two items.

Discussion

In this study we examined the risk of subsequent hip fracture among the 368 Longitudinal Study on Aging respondents whose Medicare Automated Data Retrieval System hospitalization records for 1984 through 1991 indicated that a hip fracture had occurred. Among these individuals, 27 subsequent hip fractures occurred, for a rate of 1 every 33.8 person-years. This rate is comparable to the 8% 5-year recurrence rate found in Rochester, Minn.[20] Respondents with poor perceived health and those who had trouble with dizziness had elevated risks for subsequent hip fracture. The association between poor perceived health and subsequent hip fracture mirrors that found for first hip fracture,[11] and it is consistent with a growing literature demonstrating perceived health's robust prediction of mortality and other adverse health events.[21–23] Although dizziness did not achieve statistical significance in our analysis of first hip fracture,[11] it has consistently been identified as a risk factor for hip fracture in previous studies.[1,3,12,13]

Although this report presents the first prospective analysis of the risk of subsequent hip fracture among a representative national sample of older adults, it is not without limitations. The three most important limitations are not knowing whether the subsequent hip fracture was on the same or the contralateral hip, sample size, and the length of the observation window. Only 368 respondents were at risk of having subsequent hip fracture and were at such risk for 8 years or less. Although the latter problem is not great (mean number of days to subsequent hip fracture = 613, to death without subsequent hip fracture = 674, and to censoring without subsequent hip fracture = 1132), further research on larger cohorts of those with hip fractures is needed to better characterize the risks associated with subsequent hip fracture. □

Acknowledgments

This research was supported by grant R37-AG09692 to Dr Wolinsky from the National Institutes of Health.

References

1. Melton LJ, Chrischilles EA, Cooper C, Lane AW, Riggs BL. How many women have osteoporosis? *J Bone Miner Res.* 1992;7:1005–1010.
2. Fitzgerald JF, Fagan LF, Tierney WM, Dittus RS. Changing patterns of hip fracture care before and after implementation of the prospective payment system. *JAMA.* 1987;258:218–221.
3. Cummings SR, Kelsey JL, Nevitt MC, O'Dowd KJ. Epidemiology of osteoporosis and osteoporotic fractures. *Epidemiol Rev.* 1985;7:178–208.
4. Marottoli RA, Berkman LF, Cooney LM. Decline in physical function following hip fracture. *J Am Geriatr Soc.* 1992;40:861–866.
5. Chrischilles EA, Butler CD, Davis CS, Wallace RB. A model of lifetime osteoporosis impact. *Arch Intern Med.* 1991;151:2026–2032.
6. *Healthy People 2000.* Washington, DC: Dept. f Health and Human Services; 1990.
7. Jacobsen SJ, Goldberg J, Miles TP, Brody JA, Stiers W, Rimm AA. Hip fracture incidence among the old and very old: a population-based study of 745 435 cases. *Am J Public Health.* 1990;80:871–873.
8. Jacobsen SJ, Goldberg J, Miles TP, Brody JA, Stiers W, Rimm AA. Regional variation in the incidence of hip fracture: U.S. white women 65 years of age and older. *JAMA.* 1990;264:500–502.
9. Jacobsen SJ, Goldberg J, Miles TP, Brody JA, Stiers W, Rimm AA. Seasonal variation in the incidence of hip fracture among white persons aged 65 years and older in the United States, 1984–1987. *Am J Epidemiol.* 1991;133:996–1004.
10. Jacobsen SJ, Goldberg J, Miles TP, Brody JA, Stiers W, Rimm AA. Race and sex differences in mortality following fracture of the hip. *Am J Public Health.* 1992;82:1147–1150.
11. Wolinsky FD, Fitzgerald JF. The risk of hip fracture among noninstitutionalized older adults. *J Gerontol: Soc Sci.* 1994;49. In press.
12. Melton LJ, Riggs BL. Risk factors for injury after a fall. *Clin Geriatr Med.* 1985;1:525–540.
13. Mossey JM. Social and psychological factors related to falls among the elderly. *Clin Geriatr Med.* 1985;1:541–554.
14. Fitti JE, Kovar MG. *The Supplement on Aging to the 1984 National Health Interview Survey.* Washington, DC: National Center for Health Statistics; 1987. DHHS publication PHS 87-1323.
15. Johnson RJ, Wolinsky FD. The structure of health status among older adults: disease, disability, functional limitation, and perceived health. *J Health Soc Behav.* 1993;34:105–121.
16. Harris T, Kovar MG, Suzman R, Kleinman JG, Feldman JJ. Longitudinal study of physical ability in the oldest-old. *Am J Public Health.* 1989;79:698–702.
17. Kovar MG, Fitti JE, Chyba MM. *The Longitudinal Study of Aging: 1984–1990.* Washington, DC: National Center for Health Statistics; 1992. DHHS publication PHS 92-1304.
18. Concato J, Feinstein AA, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med.* 1993;118:201–210.
19. Norusis MJ. *SPSS for Windows: Advanced Statistics.* Chicago, Ill: SPSS Inc; 1992.
20. Melton LJ, Ilstrup DM, Beckenbaugh RD, Riggs BL. Hip fracture recurrence: a population-based study. *Clin Orthop.* 1982;167:131–138.
21. Idler EL, Kasl S. Health perceptions and survival: do global evaluations of health status really predict mortality? *J Gerontol: Soc Sci.* 1991;46:S55–S65.
22. Wolinsky FD, Johnson RJ. Perceived health status and mortality among older men and women. *J Gerontol: Soc Sci.* 1992;47:S304–S312.
23. Wolinsky FD, Callahan CM, Johnson RJ. Subjective health status and mortality in the elderly. In: Vellas B, ed. *Facts and Research in Gerontology, 1994.* New York: Springer. In press.

Lifetime Milk Consumption and Bone Mineral Density in Older Women

ABSTRACT

This study examined the relation between lifetime milk consumption and both axial and appendicular bone mineral density in 581 postmenopausal White women. Positive significant, graded associations between milk consumption in adulthood and bone mineral density at the spine, total hip, trochanter, intertrochanter, and midradius, but not the ultradistal wrist or femoral neck, were observed. Adolescent milk consumption showed similar, statistically significant associations (spine and midradius). Associations were independent of age, body mass index, years postmenopausal, thiazide, estrogen and alcohol use, smoking, and exercise. Regular milk consumption in youth and adulthood is associated with better bone mineral density at cortical and trabecular sites in elderly women. (*Am J Public Health.* 1994;84:1319–1322)

Sharon Soroko, MS, Troy L. Holbrook, PhD, Sharon Edelstein, ScM, and Elizabeth Barrett-Connor, MD

Introduction

Osteoporotic low bone mineral density often leads to fractures, costly hospitalization, and prolonged use of health services.[1] Milk accounts for about half of all dietary calcium intake in North America (B. Nowlin, California Dairy Council, telephone conversation, June 1993),[2] and the importance of calcium in milk for attaining peak bone mass[3,4] and preventing osteoporosis[5–7] may relate to the quantity consumed during specific life periods.[5,8–10] In this paper, we report the association of lifetime milk consumption with bone mineral density in a community-based cohort of older women.

Methods

Study Population and Data Collection

Between 1972 and 1974, 82% of adult residents of Rancho Bernardo, Calif, enrolled in a heart disease survey.[11] From 1988 to 1991, 84% (n = 624) of surviving, local, 60- to 79-year-old women from this cohort participated in a study of bone mineral density. All were ambulatory and provided written informed consent. The 624 women were administered a standardized interview to assess their medical and life-style history, including smoking history, estrogen and thiazide use, alcohol consumption, and exercise. A standardized food-frequency questionnaire[12] was used to assess the subjects' current dietary calcium intake and their calcium supplementation history. Participants also quantified their daily milk consumption during adolescence (12 to 19 years of age), midlife (20 to 50 years of age), and older adulthood (after 50 years of age) as (1) "rarely or never" (classified as none), (2) "about every week, but not every day" (low), (3) "1 to 2 glasses per day, about every day" (medium), or (4) "3 or more glasses per day, about every meal" (high). Childhood milk intake was not queried because of expected poor recall.

Height and weight were measured with women in light clothing without shoes; body mass index (weight divided by square of height [kg/m^2]) was calculated to estimate obesity. Bone mineral density of the lumbar spine and hip was measured by dual-energy x-ray absorptiometry. Bone mineral density of the ultradistal wrist and midradius of the nondominant arm was measured by single-photon absorptiometry.[13]

This analysis includes all 581 women who provided data on milk consumption history and had bone mineral density measurements.

Statistical Analysis

SAS was used for all analyses.[14,15] The Mantel-Haenszel test was used to identify linear trends in categorical variables by milk consumption history. Analysis of covariance was used in trend analyses of continuously distributed covariates and in site-specific calculations of mean bone mineral density levels by milk consumption category; adjustments were made for the confounding effects of age, body mass index, years postmenopausal, smoking history, current use of thiazides and estrogen replacement therapy, current exercise, and alcohol consumption. Very few women reported high milk consumption during midlife (n = 27) or older adulthood (n = 17); therefore, the medium and high categories were pooled for these age periods. Separate analyses were conducted in which calcium supplementation was stratified. All statistical tests were two-tailed, with statistical significance defined as $P < .05$.

Results

The average age of participants was 70.6 years, and all were White. The women had been postmenopausal for an average of 24 years. At the time of

The authors are with the Division of Epidemiology, Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, Calif.

Requests for reprints should be sent to Elizabeth Barrett-Connor, MD, Division of Epidemiology, Department of Family and Preventive Medicine, University of California, San Diego, 9500 Gilman Dr, 0607, La Jolla, CA 92093-0607.

This paper was accepted February 4, 1994.

TABLE 1—Health Behaviors and Reported Milk Consumption at Three Age Periods: Rancho Bernardo, Calif, Women (n = 581), 1988 to 1991

	None	Low	Medium	High	*P* for Trend
Adolescence					
Consumption, %	19	17	42	23	...
Mean age, y	72	72	71	68	.0001
Mean BMI, kg/m^2	25	24	25	25	.15
Mean no. of years postmenopausal	27	25	24	22	.0001
Mean total calcium intake, mg	818	924	1068	1134	.0001
Taking estrogen, %	36	45	42	43	.40
Ever smoked, %	56	68	55	47	.05
Taking thiazides, %	24	25	25	22	.79
Exercising 3+ times per week, %	70	70	74	67	.83
Drinking alcohol 3+ days per week, %	44	49	42	41	.41
Taking calcium supplements, %	46	45	54	47	.54
Ever used calcium supplments, %	67	72	74	79	.06
Midlife					
Consumption, %	28	34	33	5	...
Mean age, y	72	71	70	66	.0001
Mean BMI, kg/m^2	24	25	25	25	.39
Mean no. of years postmenopausal	26	23	24	21	.003
Mean total calcium intake, mg	790	960	1206	1327	.0001
Taking estrogen, %	42	39	42	50	.64
Ever smoked, %	59	59	50	46	.03
Taking thiazides, %	27	28	20	14	.05
Exercising 3+ times per week, %	67	68	77	79	.03
Drinking alcohol 3+ days per week, %	47	44	39	39	.15
Taking calcium supplements, %	49	52	48	41	.52
Ever used calcium supplements, %	68	79	72	74	.49
Older adulthood					
Consumption, %	41	27	29	3	...
Mean age, y	71	71	71	67	.01
Mean BMI, kg/m^2	25	25	25	26	.40
Mean no. of years postmenopausal	25	24	23	22	.21
Mean total calcium intake, mg	760	1006	1310	1614	.0001
Taking estrogen, %	44	38	43	27	.48
Ever smoked, %	61	51	53	53	.08
Taking thiazides, %	26	29	19	7	.06
Exercising 3+ times per week, %	67	72	74	87	.05
Drinking alcohol 3+ days per week, %	48	43	39	27	.03
Taking calcium supplements, %	49	47	54	40	.57
Ever used calcium supplements, %	71	74	77	67	.27

Note. BMI = body mass index.

evaluation, 42% were using estrogen replacement therapy, 24% were using thiazides, 44% had never smoked, 71% exercised three or more times per week, and 43% drank alcohol more than three times per week. The mean total calcium intake was 1013 mg/day; 50% were currently taking calcium supplements.

As shown in Table 1, milk consumption declined with age. Only 19% of respondents reported that they drank no milk during adolescence, whereas 28% and 41%, respectively, indicated no milk consumption during midlife and older adulthood. More than 92% of women who drank little or no milk in adolescence drank little or no milk later in life. Milk intake in youth was inversely associated with smoking and positively associated with calcium supplements; no other variables were associated with teenage milk consumption. After adolescence, higher milk intake was associated with less smoking, less thiazide use, and more exercise, and after 49 years of age, higher intake was associated with less frequent alcohol consumption. Current use of estrogen or calcium supplements was not significantly associated with milk intake.

Figure 1 shows that higher milk consumption in adulthood was independently and significantly associated with higher bone mineral density levels at the midradius, spine, total hip, intertrochanter, and trochanter. Higher teenage milk intake was associated with significantly higher bone mineral density at the spine and midradius. Milk intake was not associated with bone mineral density of the ultradistal wrist. Analyses stratified by calcium supplementation revealed similar patterns (data not shown).

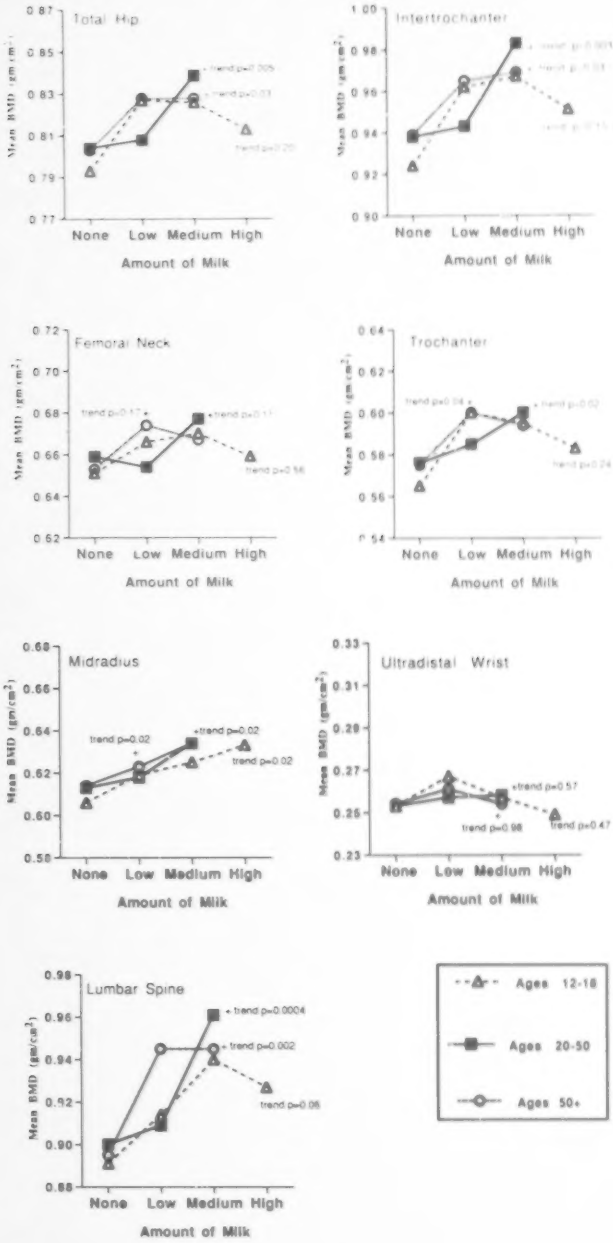
Discussion

Graded, significant associations between milk intake during three age periods and bone mineral density of the axial and appendicular skeleton were observed. These associations, independent of other major bone mineral density determinants, were strongest for milk intake during midlife (20 to 50 years of age).

The observed relation between bone mineral density and postadolescent milk consumption is consistent with evidence that a calcium-rich diet during adulthood contributes to peak bone mass[3,4,16] and/or helps minimize bone loss.[17-19] The weaker association of adolescent milk consumption with bone mineral density may reflect less variation in intake (adults have greater independence in beverage choices), inaccuracies of longer recall, or the lower per capita milk consumption during the Great Depression, which coincided with many participants' adolescent years. Alternatively, it may reflect the fact that most members of this cohort were children or teenagers when vitamin D, which mediates the absorption of calcium,[5] was first added to the US milk supply.[20]

The somewhat weaker trends for the oldest adults may reflect the downward shift in milk consumption with age. High intake after 50 years of age consequent to awareness of personal osteoporosis status did not appear to be a factor, since the relation between milk consumption and bone mineral density was not altered when the 65 women with physician-diagnosed osteoporosis or nontraumatic fractures were excluded from analysis (data not shown). Some older women may have decreased their milk consumption for health reasons.

Although self-reported milk consumption over a period of 50 years or more cannot be validated, it is unlikely that misclassification accounts for the observed association; measurement error in observational studies tends to reduce or obscure a true association.[21] Reported intake was remarkably stable over time for those who reported that they did not drink milk, with a downshift in high



Note. Means were adjusted for age, body mass index, years postmenopausal, smoking history, thiazide use, current use of estrogen replacement therapy, exercise, and alcohol consumption. Analysis of covariance was used in these calculations. BMD was measured in gm/cm². The medium and high milk consumption categories were pooled because of the small sample size of the high category for the two oldest age groups.

FIGURE 1—Multiply adjusted mean bone mineral density (BMD), by milk consumption and age group: Rancho Bernardo, Calif, women (n = 581), 1988 to 1991.

consumption with age. The proportion reporting high milk intake closely resembles that reported in similar age categories by others.[10]

Milk consumption was not associated with bone mineral density of the ultradistal wrist. Although it is possible that bone mineral density at this site is unrelated to calcium intake, a more likely explanation for this absent association is that the very low bone mineral density at the ultradistal wrist reduced the range of variation and precluded detection of differences.

Recently, Heaney, after accounting for methodological limitations in studies with negative results, argued that the literature supports a strong positive role for calcium in the prevention of age-related bone loss in women.[6] The protective effects of high dietary calcium intake on radial and spinal bone mineral density observed in this study were very similar to effects reported elsewhere.[3,10,16–18,22–24] Although we have found no other studies of milk consumption related to bone mineral density of hip sites, our findings are compatible with a previous report from this cohort in which 24-hour calcium intake was inversely related to risk of subsequent hip fracture.[19] This study lends further support to the importance of milk in the preservation of bone in women. □

Acknowledgments

This research was supported by a grant from the National Institute on Aging (AG 07181).

References

1. Holbrook TL, Grazier K, Kelsey JL, et al. *The Frequency of Occurrence, Impact, and Cost of Musculoskeletal Conditions in the United States.* Chicago, Ill: American Academy of Orthopedic Surgeons; 1985.
2. Block G, Dresser CM, Hartman AM, Carroll MD. Nutrient sources in the American diet: quantitative data from the NHANES II survey. *Am J Epidemiol.* 1985;122:13–26.
3. Halioua L, Anderson JJ. Lifetime calcium intake and physical activity habits: independent and combined effects on the radial bone of healthy premenopausal Caucasian women. *Am J Clin Nutr.* 1989;49:534–541.
4. Matkovic V, Fontana D, Tominac C, Goel P, Chestnut CH. Factors that influence peak bone mass formation: a study of calcium balance and the inheritance of bone mass in adolescent females. *Am J Clin Nutr.* 1990;52:878–888.
5. Heaney RP, Gallagher JC, Johnston CC, et al. Calcium nutrition and bone health in the elderly. *Am J Clin Nutr.* 1982;36:986–1013.
6. Heaney RP. Thinking straight about calcium. *N Engl J Med.* 1993;328:503–505.
7. Recker RR, Heaney RP. The effect of milk supplements on calcium metabolism, bone

metabolism and calcium balance. *Am J Clin Nutr.* 1985;41:254–263.
8. Barrett-Connor E. The RDA for calcium in the elderly: too little, too late. *Calcif Tissue Int.* 1989;44:303–307.
9. Cauley JA, Gutai JP, Kuller LH, et al. Endogenous estrogen levels and calcium intakes in postmenopausal women: relationship with cortical bone measures. *JAMA.* 1988;260:3150–3155.
10. Sandler RB, Slemenda CW, La Porte RE, et al. Postmenopausal bone density and milk consumption in childhood and adolescence. *Am J Clin Nutr.* 1985;42:270–274.
11. Criqui MH, Barrett-Connor E, Austin M. Differences between respondents and non-respondents in a population-based cardiovascular disease study. *Am J Epidemiol.* 1978;108:367–372.
12. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol.* 1985;122:51–65.
13. Kritz-Silverstein D, Barrett-Connor E. Grip strength and bone mineral density in older women. *J Bone Miner Res.* 1994;9:45–51.
14. *SAS Procedures Guide, Version 6.* 3rd ed. Cary, NC: SAS Institute Inc; 1990.
15. *SAS/STAT User's Guide, Version 6.* 4th ed. Cary, NC: SAS Institute Inc; 1990.
16. Picard D, Ste-Marie LG, Coutu D, et al. Premenopausal bone mineral content relates to height, weight and calcium intake during early adulthood. *Bone Miner.* 1988; 4:299–309.
17. Dawson-Hughes B, Jacques P, et al. Dietary calcium intake and bone loss from the spine in healthy postmenopausal women. *Am J Clin Nutr.* 1987;46:685–687.
18. Baran D, Sorensen A, Grimes J, et al. Dietary modification with dairy products for preventing vertebral bone loss in premenopausal women: a three-year prospective study. *J Clin Endocrinol Metab.* 1990;70: 264–270.
19. Holbrook TL, Barrett-Connor E, Wingard DL. Dietary calcium and risk of hip fracture: 14-year prospective population study. *Lancet.* 1988;2:1046–1049.
20. Jacobus CH, Holick MF, Shao Q, et al. Hypervitaminosis D associated with drinking milk. *N Engl J Med.* 1992;326:1173–1177.
21. Fleiss JL. *Statistical Methods for Rates and Proportions.* New York, NY: John Wiley & Sons Inc; 1973.
22. Bauer DC, Browner WS, Cauley JA, et al. Factors associated with appendicular bone mass in older women. *Ann Intern Med.* 1993;118:657–665.
23. Smith EL, Gilligan C, Smith PE, et al. Calcium supplementation and bone loss in middle-aged women. *Am J Clin Nutr.* 1989;50:833–842.
24. Kander B, Dempster DW, Lindsay R. Interaction of calcium nutrition and physical activity on bone mass in young women. *J Bone Miner Res.* 1988;3:145–149.

Attitudes toward Body Size and Dieting: Differences between Elderly Black and White Women

June Stevens, PhD, MS, Shiriki K. Kumanyika, PhD, MPH, and Julian E. Keil, DrPH

ABSTRACT

Eating restraint and body size perceptions of 404 White and African-American women 66 to 105 years of age (mean age = 73 years) were assessed by questionnaire. Compared with overweight White women, overweight Black women were 0.6 times as likely to feel guilty after overeating, 0.4 times as likely to diet, 2.5 times as likely to be satisfied with their weight, and 2.7 times as likely to consider themselves attractive. Among those who were not overweight, Black women were half as likely as White women to consider themselves overweight. Compared with Black women, White women perceived themselves to be larger and reported a lower ideal body weight. (*Am J Public Health.* 1994;84: 1322–1325)

Introduction

The prevalence of obesity in African-American women is approximately twice that in White women.[1,2] Although the causes of this high Black–White ratio remain unclear, differences between African-American and White women in the motivation or ability to lose weight are probable contributing factors.[3–6]

Several researchers have reported differences in perceived or ideal body size in adolescent and college-aged African-American and White females.[5,7,8] Adult African-American women have also been shown to be less likely to think that being overweight is a problem[9] or to perceive themselves as overweight.[10] National data on the prevalence of weight loss attempts indicate that dieting patterns are similar among overweight African-American and White adult women and that, among women who are not overweight, dieting is less common in African Americans than in Whites.[10–14]

The purpose of this research was to examine attitudes toward eating and body size perceptions in elderly Black and White women. Reports of body size perceptions and eating restraint in elderly women are rare.[14] These attitudes, however, are relevant to older people in that they may influence the effectiveness of dietary interventions, including advice to lose or gain weight. Also, it is of interest to document attitudes toward eating and body size perceptions in Black and White women born at the turn of the century because there has been a secular change in societal pressures for thinness in American women.[15–18]

At the time this paper was written, June Stevens and Julian E. Keil were with the Department of Biometry and Epidemiology, Medical University of South Carolina, Charleston. June Stevens is now with the School of Public Health, University of North Carolina, Chapel Hill. Shiriki K. Kumanyika is with the Center for Biostatistics and Epidemiology, College of Medicine, Pennsylvania State University, Hershey.

Requests for reprints should be sent to June Stevens, PhD, MS, Departments of Epidemiology and Nutrition, School of Public Health, University of North Carolina, Chapel Hill, NC 27599.

This paper was accepted March 16, 1994.

TABLE 1—Eating Restraint and Body Image in Elderly Women

	Women Providing Affirmative Response, %				Adjusted Odds Ratios (95% Confidence Interval)			
	Black		White		Black		White	
	Not Overweight	Over-weight	Not Overweight	Over-weight	Not Overweight	Over-weight	Not Overweight	Over-weight
Eating restraint								
Ever eaten less than you wanted to keep weight down	27*	68	41*	69	1.0 (0.6, 1.5)	2.2 (1.5, 3.1)	1.0 . . .	1.9 (1.4, 2.5)
Feel guilty after overeating	0*	21	9*	37	0.5 (0.2, 1.7)	1.5 (0.9, 2.3)	1.0 . . .	2.5 (1.8, 3.5)
Ever dieted to lose weight	51	54	49*	66	1.2 (0.6, 2.6)	1.2 (0.8, 1.8)	1.0 . . .	2.7 (1.7, 4.2)
Body image								
As adult, tried to gain weight	16*	5	22*	10	0.7 (0.4, 1.2)	0.35 (0.2, 0.6)	1.0 . . .	0.6 (0.4, 0.9)
Believe that extra weight at older ages is good	10	9	23	30	0.5 (0.3, 1.0)	0.6 (0.4, 1.0)	1.0 . . .	1.2 (0.9, 1.6)
Satisfied with current weight	76*	58	70*	26	1.0 (0.7, 1.6)	0.6 (0.4, 0.9)	1.0 . . .	0.4 (0.3, 0.5)
Consider yourself over-weight	7*	70	41*	94	0.5 (0.2, 0.9)	2.0 (1.4, 2.9)	1.0 . . .	4.5 (3.0, 6.9)
Spouse/significant other thinks you are over-weight[a]	0*	46	19*	62	0.7 (0.2, 2.2)	2.5 (1.3, 4.9)	1.0 . . .	2.7 (1.7, 4.3)
Rate your figure somewhat attractive or attractive	73	70	67*	37	1.2 (0.8, 2.0)	1.0 (0.7, 1.4)	1.0 . . .	0.5 (0.4, 0.7)

Note. Odds ratios and 95% confidence intervals are adjusted for age and education. White women who were not overweight were used as the reference group.

[a]Responses used for this estimate included 25 not overweight Black women, 44 overweight Black women, 56 not overweight White women, and 64 overweight White women.

*$P < .05$ (within race groups).

Methods

Participants in the Charleston Heart Study constitute a probability sample, recruited in 1960, of residents of Charleston County, South Carolina, over the age of 65. The data used in the present study were taken from the 1991 reinterview of 517 women from 66 to 105 years of age in the Charleston Heart Study.[19] Complete nutrition information was collected from 404 women (278 White and 126 Black). The mean age of these 404 women was 3 years younger than that of the women for whom only incomplete data were available (73 years vs 76 years; $P < .001$).

A questionnaire was developed and tested to examine attitudes toward eating, body size perceptions, and dieting practices. Some items were adapted from other surveys.[5,20-22] Line drawings of women[20] ranging from very thin to obese (scored 1 to 9) were used to measure perceptions of actual and ideal body size. A body size discrepancy score was calculated as the difference between the perceived actual and ideal size.

All items were pretested by means of in-depth interviews. Test–retest reliability was determined in 50 subjects older than 65 years of age over a period of 3 to 6 weeks. The mean kappa values were 0.6 for items concerning attitudes toward eating and 0.7 for questions on body size perceptions, indicating good reproducibility.[23]

A body mass index of 27.3 kg/m^2 was used as a cutpoint to indicate overweight.[24] Women with body mass indexes of less than 27.3 kg/m^2 are referred to here as "not overweight."

Statistical analyses included the two-sample t test, analysis of variance, chi-square tests, and multivariate logistic regression. A frequency count of 1 was

TABLE 2—Odds Ratios for Eating Restraint and Body Image in Overweight Black vs Overweight White Women

	Odds Ratio	95% Confidence Interval
Eating restraint		
Ever eaten less than you wanted to keep weight down	0.9	0.7, 1.4
Feel guilty after overeating	0.6	0.4, 0.9
Ever dieted to lose weight	0.4	0.2, 0.7
Body image		
As adult, tried to gain weight	0.8	0.4, 1.6
Believe that extra weight at older ages is good	0.6	0.4, 0.9
Satisfied with current weight	2.5	1.7, 3.7
Consider yourself overweight	0.3	0.1, 0.5
Spouse/significant other thinks you are overweight[a]	0.9	0.5, 2.0
Rate your figure somewhat attractive or attractive	2.7	1.7, 4.1

Note. Odds ratios and 95% confidence intervals are adjusted for body mass index, age, and education. Overweight White women were used as the reference group.

[a]Responses used for this estimate included 44 overweight Black women and 64 overweight White women.

TABLE 3—Perceived and Ideal Body Size in Elderly Women

	Black	White	P^a	Not Overweight	Over-weight	P^b
Silhouette most like you[c]	3.9	4.8	.001	3.7	5.4	.001
Ideal silhouette[c]	3.6	3.5	.28	3.4	3.8	.001
Discrepancy score[d]	.03	1.4	.001	0.3	1.6	001
Reported ideal weight, kg	65	62	.001	60.0	68.1	001
Calculated ideal body mass index, kg/m^2	26.0	24.1	.001	22.6	27 5	001
Difference between actual and ideal body mass index, kg/m^2	1.0	2.8	.001	0.2	4.7	.001

[a]Corrected means and *P* values from model including age, education, and body mass index.
[b]Corrected means and *P* values from model including age, education, and race.
[c]On a scale of 1 to 9.
[d]Difference between perceived and ideal.

given to empty cells for the estimate of odds ratios.

Results

The mean body mass index was 23 kg/m^2 in Black and White women who were not overweight, but overweight Black women were an average of one body mass index unit heavier than overweight White women (33 kg/m^2 vs 32 kg/m^2; $P < .01$). Sixty-five percent of the Black women and 35% of the White women were in the overweight category. Four percent of both Black and White women were underweight (i.e., had a body mass index less than or equal to 19.1 kg/m^2). Black women had fewer years of education than White women (7 years vs 11 years; $P < .001$). Educational attainment did not differ between the weight categories within racial groups.

Tables 1 and 2 show responses to nine items for Black and White women within weight categories. Odds ratios comparing the responses of women in different weight status and race categories are shown. In both racial groups, overweight women reported more dieting and less satisfaction with their body size.

Table 3 compares perceptions and ideals of body size by race and weight status. The overall effects of race and weight status were statistically significant for all but one comparison. Responses of African-American women indicated that they were more satisfied with their body size than were White women. Responses of overweight women indicated that their ideal body mass index was higher by 4.9 kg/m^2 than that of women who were not overweight. In fact, the mean ideal body mass index calculated for overweight Black women—28.7 kg/m^2—would be categorized as overweight using the National Center for Health Statistics cut-point of 27.3 kg/m^2. Among African-American women, the percentage who considered their figures attractive was unrelated to the percentage who were overweight.

Discussion

In this study, body size perceptions and attitudes toward eating differed by both weight status and race in an elderly cohort. The differences between women who were and were not overweight agree with those found in previous studies of exclusively or predominantly young women, which have shown that obese women (both Black and White) are more likely than thin women to report eating restraint.[5,25,26] The differences between African-American and White women in body size perceptions were striking and suggest underlying differences in cultural attitudes toward overweight in this cohort.

Insights into the possible nature of this difference in cultural attitudes come from nutritional anthropological studies of attitudes toward obesity[15,27] and studies of socioeconomic factors in obesity.[28,29] Positive attitudes toward being moderately overweight are characteristic of economically disadvantaged societies in which thinness is a sign of poor health. Theoretically, differences in body weight between Black and White women will be reduced as African Americans gain greater social and economic advantages. However, movement of African-American women toward the attitudinal profiles of White women is not necessarily desirable. Black women who are not overweight are less likely to have an inappropriate body image. The challenge to the public health community is to encourage weight reduction when appropriate without forcing on Black women a negative self-image and a neurotic preoccupation with weight. □

Acknowledgments

This work was funded by grants RO1-HL-42305 and HL-31392O8 from the National Institutes of Health.

We acknowledge the contribution to this work made by the late Dr Edwin Boyle, Jr, who initiated the Charleston Heart Study. We also thank Laurie Graybeal for her assistance in the collection of repeatability information on the questionnaire.

References

1. *Health United States 1990.* Washington, DC: National Center for Health Statistics; 1991. DHHS publication PHS 91-1232.
2. Gillum R. Overweight and obesity in black women: a review of published data from the National Center for Health Statistics. *J Natl Med Assoc.* 1987;79:865–871.
3. Kumanyika S. Obesity in black women. *Epidemiol Rev.* 1987;9:31–50.
4. Kumanyika S, Obarzanek E, Stevens V, Hebert P, Whelton P. Weight-loss experience of black and white participants in NHLBI-sponsored clinical trials. *Am J Clin Nutr.* 1991;53:1631S–1638S.
5. Kumanyika S, Wilson J, Guilford-Davenport M. Weight-related attitudes and behaviors of black women. *J Am Diet Assoc.* 1993;93:416–422.
6. Kahn H, Williamson D, Stevens J. Race and weight change in US women: the roles of socioeconomic and marital status. *Am J Public Health.* 1991;81:319–323.
7. Desmond S, Price J, Hallinan C, Smith D. Black and white adolescents' perceptions of their weight. *J Sch Health.* 1989;59:353–358.
8. Rucker C, Cash T. Body images, body size perceptions, and eating behaviors among African-American and white college women. *Int J Eating Disord.* 1991;12:291–299.
9. Rand C, Kuldau J. Epidemiology of bulimia and symptoms in a general population: sex, age, race, and socioeconomic status. *Int J Eating Disord.* 1992;11:37–44.
10. Dawson D. Ethnic differences in female overweight: data from the 1985 National Health Interview Survey. *Am J Public Health.* 1988;78:1326–1328.
11. Williamson D, Serdula M, Anda R, Levy A, Byers T. Weight loss attempts in adults: goals, duration, and rate of weight loss. *Am J Public Health.* 1992;82:1251–1257.
12. Bennett E. Weight-loss practices of overweight adults. *Am J Clin Nutr.* 1991;53:1519S–1521S.
13. Stephenson M, Levy A, Sass N, McGarvey W. 1985 NHIS findings: nutrition knowledge and baseline data for the weight-loss objectives. *Natl Health Rep.* 1987;102:61–67.
14. *Health Promotion and Disease Prevention.*

Washington, DC: National Center for Health Statistics; 1988. DHHS publication PHS 88-1591.
15. Brown P, Konner M. An anthropological perspective on obesity. *Ann NY Acad Sci.* 1987;499:29–46.
16. Garner D, Garfinkel P, Schwartz D, Thompson M. Cultural expectations of thinness in women. *Psychol Rep.* 1980;47:483–491.
17. Wiseman C, Gray J, Mosimann J, Ahrens A. Cultural expectations of thinness in women. An update. *Int J Eating Disord.* 1992;11:85–89.
18. Striegel-Moore R, Silberstein L, Rodin J. Toward an understanding of risk factors for bulimia. *Am Psychol.* 1986;41:246–263.
19. Boyle E. Biological patterns in hypertension by race, sex, body weight and skin color. *JAMA.* 1970;213:1637–1643.
20. Stunkard A, Sorensen T, Schulsinger F. Use of the Danish adoption register for the study of obesity and thinness. In: Kety S, Rowland L, Sidman R, Matthysse S, eds. *Genetics of Neurological and Psychiatric Disorders.* New York, NY: Raven Press; 1983.
21. Stunkard A. "Restrained eating": what it is and a new scale to measure it. In: Ciofti L et al., eds. *The Body Weight Regulatory System: Normal and Disturbed Mechanisms.* New York, NY: Raven Press; 1981:243–251.
22. Herman C, Polivy J. Anxiety, restraint, and eating behavior. *J Abnorm Psychol.* 1975;84:666–672.
23. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
24. Sichieri R, Everhart J, Hubbard V. Relative weight classifications in the assessment of undeweight and overweight in the United States. *Int J Obes.* 1992;16:303–312.
25. Hibscher J, Herman C. Obesity, dieting, and the expression of "obese" characteristics. *J Comp Physiol Psychol.* 1977;91:374–380.
26. Johnson W, Lake L, Mahan J. Restrained eating: measuring an elusive construct. *Addict Behav.* 1983;8:413–418.
27. Ritenbaugh C. Obesity as a culture-bound syndrome. *Cult Med Psychiatry.* 1982;6:347–361.
28. Goldblatt P, Moore M, Stunkard A. Social factors in obesity. *JAMA.* 1965;192:1039–1044.
29. Sobal J, Stunkard A. Socioeconomic status and obesity: a review of the literature. *Psychol Bull.* 1989;105:260–275.

The 1966 Enactment of Medicare: Its Effect on Discharges from Los Angeles County–Operated Hospitals

Peter A. Glassman, MD, MSc, Robert M. Bell, PhD, and Robert E. Tranquada, MD

ABSTRACT

The effect of Medicare on two public hospitals in Los Angeles County was analyzed by examining the percentage of patients 65 years of age and older among all discharges from 1958 through 1971. At Harbor General Hospital, discharges of elderly patients had dropped from 21.7% to 7.9% by late 1966; at Los Angeles County General Hospital, discharges decreased from 15.3% to 10.7% between 1966 and 1967. Monitoring public hospitals' demographic changes after enacting a national health plan may provide information on patients' and providers' acceptance of insurance and on resources needed by public hospitals to care for those left without coverage. (*Am J Public Health.* 1994;84:1325–1327)

Introduction

A major debate shaping US health care policy centers on the optimal mix of public and private insurance needed to finance care for nearly 37 million uninsured Americans. Proposed legislation includes employer-based insurance, single payer universal coverage, and expansion of Medicaid. There is controversy about the effects that any of these options would have on those currently uninsured.

In considering the potential effects of health reforms, it is important to examine the role of public hospitals, which currently serve a disproportionate share of the uninsured and underinsured. In this regard, an instructive chain of events began on July 1, 1966, when Medicare began providing health insurance for many previously uninsured elderly people. Prior to Medicare, an estimated 56% to 75% of the nation's elderly had inadequate health insurance, and many depended on public hospitals for medical care.[1–4]

Information on how Medicare has affected public hospitals is somewhat sparse.[5,6] In New York City, indirect information indicates that, by 1975, a large proportion of Medicaid and Medicare recipients were receiving care in the private sector. The exact timing and the magnitude of the change from public to private care are not clear, however.[7]

This study examined the effect of Medicare enactment on discharges of elderly patients (those 65 years of age and older) from the two county-operated acute hospitals in Los Angeles County. We hypothesized that health insurance for uninsured and underinsured elderly people produced a rapid and permanent shift of these patients into the private sector.

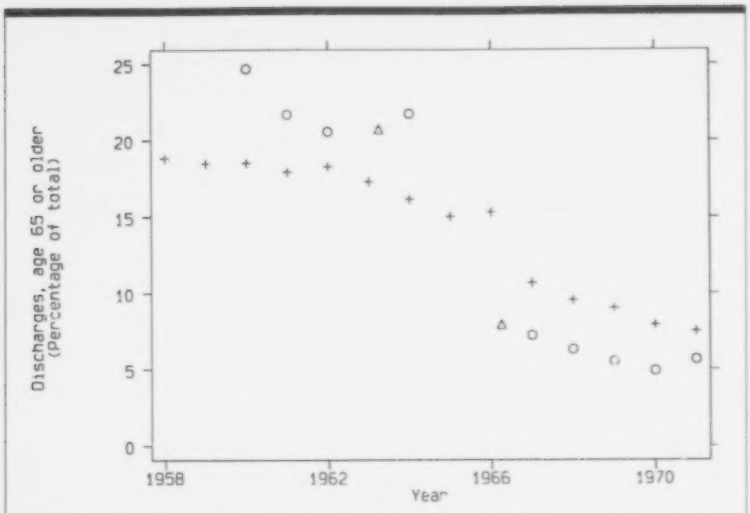
Methods

Data were collected from the stored records of Los Angeles County General Hospital (now Los Angeles County/

The authors are with the RAND Corporation, Santa Monica, Calif. Peter A. Glassman is also with the West Los Angeles Veterans Affairs Medical Center. Robert E. Tranquada is also with the Schools of Medicine and Public Administration, University of Southern California, Los Angeles.

Requests for reprints should be sent to Peter A. Glassman, MD, MSc, Division of General Internal Medicine (691/W111G), VA Medical Center, West Los Angeles, Wilshire and Sawtelle Blvds, Los Angeles, CA 90073.

This paper was accepted December 23, 1993.



Note. ○ = Harbor General Hospital; △ = Harbor General Hospital, July to December only; + = Los Angeles County General Hospital.

FIGURE 1—Percentage of elderly discharges: Los Angeles County General Hospital and Harbor General Hospital, 1958 through 1971.

University of Southern California Medical Center) and Harbor General Hospital (now Los Angeles County Harbor/University of California at Los Angeles Medical Center). We computed the percentage of elderly patients among all discharges (including deaths) for the years 1958 through 1971. Data were not available for Harbor General Hospital for 1958, 1959, 1965, and the first 6 months of 1963 and 1966.

We conducted time series analyses for each hospital to determine whether (1) there was a trend in the proportion of discharges of elderly patients prior to Medicare and (2) the Medicare program altered the trend.

To separate a possible Medicare-induced drop from an overall downward trend in discharges from 1958 to 1971, we fit the Los Angeles County General Hospital data to a linear regression model that allowed for distinct slopes during the periods 1958 through 1962, 1963 through 1966, and 1967 through 1971 and allowed for shifts at 1963 and 1967. (We did not do so for Harbor General because of the missing data.) This allowed evaluation of the average percentage change in discharges per year during the three time periods (linear trends). Two-tailed t tests were used in determining significant regression coefficients.

Results

For both hospitals, there was a downward trend in the percentage of elderly discharges beginning in 1963 and continuing up until Medicare enactment (Figure 1). At Harbor General, a dramatic drop from 21.7% to 7.9% occurred between the end of 1964 and 1966. At County General, the drop from 15.3% to 10.7% appears to have occurred later, between 1966 and 1967. The shift at County General, after controlling for the pre- and post-Medicare trends, was 3.55% ($t = -7.9$, $P < .001$); the shift at 1963 was not statistically significant. The slopes of the downward trends from 1958 to 1962 and from 1967 to 1971 were −0.7% ($t = -4.7$, $P = .002$) and −0.8% per year ($t = -7.4$, $P < .001$), respectively. The pre- and post-Medicare slopes did not differ significantly from each other. By 1971, the proportions of elderly discharges at County General and Harbor General had fallen to 7.5% and 5.6% of total discharges, respectively.

Discussion and Conclusions

Medicare enactment was associated with a large drop in the number of elderly patients being cared for in Los Angeles' county-operated hospitals, even as the population of those 65 years of age and older in Los Angeles County increased by nearly 100 000 during the 1960s.[8,9] Elderly patients currently account for about 4% of total discharges from Los Angeles County hospitals, indicating that the effect of Medicare was not a temporary phenomenon.[10]

The downward trend occurring before Medicare at Los Angeles County General may be due to the Kerr-Mills legislation implemented in California in 1962. This legislation allowed some indigent elderly patients to receive care in private settings but, because of its limited coverage, would not likely explain the large patient shift seen during the period from 1965 to 1967.

Not all elderly patients left the county-operated hospital system after the implementation of Medicare. Possible reasons include the following: (1) County hospitals provided the nearest comprehensive care facility; (2) some patients were ineligible for Medicare; (3) some patients voluntarily continued their care at a county hospital; and (4) some recipients received emergency inpatient care at county hospitals. Lack of adequate local private providers also could have been a factor. A combination of these elements may explain the smaller initial decrease seen at County General, which served a larger and poorer section of Los Angeles than did Harbor General. Other barriers to obtaining care, such as transportation and language difficulties, may also have affected accessibility to private services.

As mentioned earlier, a significant movement of elderly patients away from Los Angeles County hospitals occurred after the implementation of Medicare. Although confirmatory data from private institutions are unavailable, it appears that many newly insured elderly people quickly opted for care at private facilities and offices where they perceived that a higher quality of care was available.

Medicare enactment thus caused a differential impact on the private and public sectors, a scenario that will probably be repeated if proposed reforms do not succeed in implementing health coverage for all. Those who do receive coverage (i.e., those who are employed but not insured), assuming that private providers accept it, will probably be absorbed into fee-for-service or managed care. Public hospitals will be left to provide care for the remainder of the population: those who are unemployed, part-time laborers, those who are ineligible, and immigrants. The way in which public hospitals adapt to the change in case mix, funding, and

patient needs may be as important to the success of health reform as the overall reform package itself. Public hospitals will probably need assistance in defining their role and function in a system of managed competition. They will need to be monitored and, if necessary, allocated new resources to ensure that they will be able to continue offering appropriate care to those in need. □

Acknowledgments

We would like to thank Robert H. Brook, MD, ScD, Arleen Leibowitz, PhD, and Howard Freeman, PhD, for their comments and assistance.

References

1. Gornick M, Greenberg JN, Eggers PW, Dobson A. Twenty years of Medicare and Medicaid: covered populations, use of benefits, and program expenditures. *Health Care Finance Rev.* 1985;6(suppl):13–59.
2. Mariano LA. Growth of the Medicare population. *Health Care Finance Rev.* 1989; 10:123–124.
3. Fein R. *Medical Care, Medical Costs: The Search for a Health Insurance Policy.* Cambridge, Mass: Harvard University Press; 1986:52–67.
4. Williams SJ, Torrens PR, eds. *Introduction to Health Services.* 3rd ed. Albany, NY: Delmar Inc; 1988:335–370.
5. Coe RM, Brehm HP, Peterson WA. Impact of Medicare on the organization of community health resources. *Milbank Mem Fund Q Health Soc.* 1974;52(3):231–264.
6. Gornick M. Trends and regional variations in hospital use under Medicare. *Health Care Finance Rev.* 1982;3:41–73.
7. Piore N, Lieberman P, Linnane J. Public expenditures and private control? Health care dilemmas in New York City. *Milbank Mem Fund Q Health Soc.* 1977;55(1):79–116.
8. *County and City Data Book, 1967.* Washington, DC: US Bureau of the Census; 1967:32.
9. *County and City Data Book, 1972.* Washington, DC: US Bureau of the Census; 1972:54.
10. *Quarterly Report of Hospitals, Fourth Quarter, 1991.* Sacramento, Calif: Office of the Statewide Health Planning and Development, California Department of Health Services; 1991.

Sociodemographic and Health Factors in Driving Patterns after 50 Years of Age

Raynard Kington, MD, PhD, David Reuben, MD, Jeannette Rogowski, PhD, and Lee Lillard, PhD

ABSTRACT

Data from the 1990 Panel Study of Income Dynamics were used to predict, by means of logistic regression, the likelihood that people who had previously driven would continue to drive and to drive after dark after 50 years of age. The results support the conclusion that driving patterns appear to be explained partly by a combination of sociodemographic factors and health status. Furthermore, it is shown that those reported to drive for nondrivers appear to be the same individuals known to provide most informal support for functionally impaired elderly people. (*Am J Public Health.* 1994;84:1327–1329)

Introduction

As the ranks of older Americans grow, more older people will be driving. Using a population-based study sample that included a substantial representation of older African Americans, we attempted to answer the following questions: What sociodemographic, economic, and health factors are associated with driving and driving after dark among older people? Among older people who do not drive, who assumes responsibility for driving them?

Methods

This study was based on reports of driving habits from a 1990 health supplement questionnaire mailed to household heads and spouses more than 50 years of age taking part in the nationally representative Panel Study of Income Dynamics.[1] Of the 3277 individual eligible household heads and spouses, 2429 (74%) completed surveys.

Sociodemographic variables studied in this analysis included age, sex, race, education, marital status, gross income of head of household and spouse (in 1988), number of adults in the household, urbanicity of the county of residence, and region of the country. Health status was measured by three multi-item subscales based on the RAND SF 36 health survey: general health perceptions and physical and emotional role functioning (higher scores indicate better health status for each index).[2] Functional status was measured by reported limitations in the following activities of daily living and instrumental activities of daily living: bathing, dressing, toileting, transferring, walking inside the home, eating, doing light housework, preparing meals, doing laundry, shopping for groceries, managing money, taking medications, and telephoning. Also, the presence of the following self-reported chronic medical conditions was noted: arthritis, hypertension, hearing impairment, visual impairment, con-

Raynard Kington, Jeannette Rogowski, and Lee Lillard are with the RAND Corporation, Santa Monica, Calif. Raynard Kington and David Reuben are with the UCLA School of Medicine, Multicampus Program in Geriatric Medicine and Gerontology, Los Angeles, Calif.

Requests for reprints should be sent to Raynard Kington, MD, PhD, RAND, 1700 Main St, PO Box 2138, Santa Monica, CA 90407-2138.

This paper was accepted February 4, 1994.

TABLE 1—Demographic, Health, and Functional Status Data of Study Participants

	Sample (n = 2429), %
Age, y	
50–59	34
60–69	36
70–79	21
80–89	8
90+	1
Sex	
Male	41
Female	59
Race/ethnicity	
White	72
African American	24
Other	2
Marital status	
Married	63
Never married	3
Widowed	22
Divorced/separated	11
General health	
Excellent, very good, or good	65
Fair or poor	35
Functional status	
Limitation in ≥ 1 activity of daily living	17
Limitation in ≥ 1 instrumental activity of daily living	21
Medical conditions	72
≥ 1 chronic medical condition	

TABLE 2—Predictors of Older Persons Continuing to Drive after 50 Years of Age and Driving after Dark

	Adjusted Odds Ratio[a] (95% Confidence Interval)	
	Driving after 50 Years of Age (n = 1880)	Driving after Dark (n = 1536)
60–70 years of age	0.07 (0.02, 0.03)	0.36 (0.18, 0.71)
70–80 years of age	0.03 (0.006, 0.012)	0.09 (0.05, 0.19)
80–90 years of age	0.004 (0.001, 0.019)	0.03 (0.01, 0.07)
90+ years of age	0.002 (0.0002, 0.014)	0.02 (0.002, 0.21)
Female	0.30 (0.17, 0.53)	0.17 (0.10, 0.29)
Married	3.06 (1.59, 5.92)	. . .
Education > 12 y	2.24 (1.21, 4.17)	3.39 (1.95, 5.92)
No. of adults in household	0.64 (0.42, 0.98)	. . .
Urban county of residence, %	0.99 (0.98, 1.00)	. . .
West	2.76 (1.30, 5.88)	. . .
North Central	2.33 (1.16, 4.65)	. . .
Health perceptions index[b]	1.18 (1.10, 1.27)	. . .
Emotional role functioning index[c]	0.86 (0.64, 1.17)	1.39 (1.06, 1.82)
Functional limitation in shopping	. . .	0.30 (0.10, 0.90)
Functional limitation in preparing meals	6.45 (1.87, 22.22)	. . .
Functional limitation in taking medication	0.07 (0.01, 0.45)	. . .
Functional limitation in using the phone	1.20 (0.34, 4.23)	17.32 (1.00, 300.67)
Visual impairment	0.47 (0.27, 0.81)	0.27 (0.16, 0.47)
Arthritis	3.08 (1.82, 5.23)	1.76 (1.08, 2.86)
Diabetes mellitus	0.84 (0.39, 1.83)	0.34 (0.18, 0.66)
Major neurological impairment	0.10 (0.04, 0.25)	0.26, (0.08, 0.80)
Pseudo R^2	.5113	.3464

Note. Race, geographic region (South [omitted category: Northeast]), number of children in household, 1988 household income, dummy variables for activities of daily living and instrumental activities of daily living, an index for physical role functioning, and dummy variables for congestive heart failure, hypertension, hearing impairment, cancer, and angina were also controlled. Only adjusted odds ratios for significant coefficients in either equation are presented. The age category 50–59 years was omitted.

[a]Estimated by multivariate logistic regression.

[b]Adjusted odds ratio for an index that takes on values from 5–25 (higher score means better health) (thus, the adjusted odds ratio for a one-unit change in the index).

[c]Adjusted odds ratio for an index that takes on values from 0–3 (higher score means better health).

gestive heart disease, diabetes mellitus, angina, and major neurological impairment. Respondents were asked whether they drove and, if they did, whether they drove after dark. If they did not drive, they were asked whether they had ever driven and, if so, when they had stopped driving and their reasons for doing so.

Logistic regression was used to estimate the likelihood of a respondent reporting driving in 1989 and driving after dark. People who had never driven or who had stopped driving before 50 years of age were excluded from these analyses. Individual observations in the regressions were weighted by a function of the probability of being included in the original Panel Study of Income Dynamics sample, continuing in the follow-up sample through 1990, and responding to the mail-out survey.

Results

Table 1 describes the study population. Of the 2429 respondents, 1716 (71%) reported driving in 1989, and 1442 (84%) of these individuals reported driving after dark. Of those 577 who did not drive, about half (284) had never driven. The average age at which former drivers had stopped driving was 60 years. The most common reasons for stopping were a health problem other than eyesight or hearing impairment (30%), "trouble with eyesight" (29%), and "not comfortable driving" (27%). Only 4% reported having their driver's license revoked. Eighteen percent and 15%, respectively, reported that they could not afford a car and the accompanying insurance premiums. For those who did not drive, the sources of transportation were an adult child (39%), a spouse (21%), a sibling (7%), other relatives (24%), a friend (23%), and taxi drivers or other paid individuals (15%).

Table 2 presents significant coefficients ($P \leq .05$) for logistic regressions estimating the likelihood that an individual reported driving or driving after dark in 1989. Older people, women, and those in more urban counties and in households with more adults were less likely to drive, and better educated and married individuals and residents of the West and North Central regions were more likely to drive. Individuals with better self-reported general health were more likely to drive, but physical and emotional role functioning were not significant predictors of driving. The only activities of daily living or instrumental activities of daily living with significant

coefficients were "difficulty with taking medications" and "difficulty with preparing meals," and those with the latter condition were more likely to drive. Of the medical conditions, individuals with major neurological conditions and visual impairment were significantly less likely to drive, and those with arthritis were more likely to drive. The pattern of predictors of not driving after dark was similar to that for the predictors of not driving. The self-perceived health index, marital status, the number of adults in the household, and the regional dummies, however, were not significant predictors of driving after dark. People with diabetes and with worse scores on the emotional role functioning index were more likely to restrict driving after dark.

Discussion

The patterns of association between driving and sociodemographic factors were similar to those found in previous research.[3-5] Of note was the new finding that individuals who lived in households with more adults were less likely to drive. This pattern may reflect the choice of elderly people to stop driving when other adults are readily available to drive or the decision of those who can no longer drive to choose living arrangements that provide them a larger pool of potential drivers on which to draw.

Multivariate analyses of the contribution of health status in explaining driving patterns revealed that few of the self-reported functional and health status measures were predictors of driving patterns. Instead, a combination of the sociodemographic variables, an index of self-perceived health, and three medical conditions—visual impairment, arthritis, and major neurological impairment—explained much of the variation.

Surprisingly, the presence of arthritis was associated with a higher probability of continuing driving. Perhaps other modes of transportation (e.g., most buses) are so poorly suited to accommodate those with functional limitations due to chronic arthritis that they continue driving as the least difficult means of reliable transportation.

Restricting driving after dark is often an intermediate step in self-regulation of driving before complete discontinuation. In the equations for predicting driving after dark, the presence of diabetes, which may lead to visual impairment, also led to a greater likelihood of restricting such driving.

Driving patterns appear to be explained partly by a combination of sociodemographic factors and health status. Our results go beyond previous research in defining more precisely various dimensions of health and socioeconomic status and in controlling for regional characteristics such as urbanicity. Furthermore, the results show that those reported to drive for nondrivers appear to be the same individuals known to provide most informal support for functionally impaired elderly people.[6] □

Acknowledgments

This study was supported by National Institute on Aging program grant P01-AG08291-04. Dr Kington was also supported by a junior faculty development award from the UCLA/National Institute on Aging Pepper Older American Independence Center (grant P60-AG10415).

References

1. Becketti S, Gould W, Lillard L, Welch F. The Panel Study of Income Dynamics after fourteen years: an evaluation. *J Labor Economics.* 1988;6:472–492.
2. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Economics.* 1993;2:217–227.
3. Jette A, Branch L. A ten-year follow-up of driving patterns among the community-dwelling elderly. *Hum Factors.* 1992;34:25–32.
4. Rosenbloom S. The mobility needs of the elderly. In: *Transportation in an Aging Society.* Washington, DC: National Research Council; 1988.
5. Campbell MK, Bush TL, Hale WE. Medical conditions associated with driving cessation in community-dwelling, ambulatory elders. *J Gerontol.* 1993;48:S230–S234.
6. Doty P. Family care of the elderly: the role of public policy. *Milbank Q.* 1986;64:34–75.

A2623-1

American Journal of Public Health

September 1994, Vol. 84, No. 9 Established 1911

ED KIANKA
HEALTH REPORTER
CONGRESSIONAL INFORMATION SVC
4520 EAST-WEST HIGHWAY
BETHESDA MD 20814
9375

American Public Health Association

Featuring Perinatal Health

Editorials

1369 The Future of Community-Based Cardiovascular Disease Intervention Studies *M. A. Winkleby*
1372 The Impressionable Fetus? Fetal Life and Adult Health *N. Paneth*
1374 What You Measure Is What You Get: Prenatal Care and Women's Health *P. H. Wise*
1376 Estimating the Costs of Pregnancy-Related Care *M. C. McCormick and S. K. Lee*
1377 More Evidence for the Public Health Value of Family Planning *L. V. Klerman and J. A. Klerman*

1993 Presidential Address

1379 Women Are Organizing: Environmental and Population Policies Will Never Be the Same *H. Rodriguez-Trias*

Articles

VARIOUS TOPICS

1383 Community Education for Cardiovascular Disease Prevention: Risk Factor Changes in the Minnesota Heart Health Program *R. V. Luepker, D. M. Murray, D. R. Jacobs, Jr, M. B. Mittelmark, N. Bracht, R. Carlaw, R. Crow, P. Elmer, J. Finnegan, A. R. Folsom, R. Grimm, P. J. Hannan, R. Jeffrey, H. Lando, P. McGovern, R. Mullis, C. L. Perry, T. Pechacek, P. Pirie, J. M. Sprafka, R. Weisbrod, and H. Blackburn*
1394 How Effective Is Drug Abuse Resistance Education? A Meta-Analysis of Project DARE Outcome Evaluations *S. T. Ennett, N. S. Tobler, C. L. Ringwalt, and R. L. Flewelling*
1402 Drinking Behavior in Relation to Cause of Death among US Adults *G. Li, G. S. Smith, and S. P. Baker*

PERINATAL HEALTH

1407 Maternal Smoking during Pregnancy and Smoking by Adolescent Daughters *D. B. Kandel, P. Wu, and M. Davies*
1414 An Evaluation of the Kessner Adequacy of Prenatal Care Index and a Proposed Adequacy of Prenatal Care Utilization Index *M. Kotelchuck*
1421 Prenatal Alcohol Exposure, Birthweight, and Measures of Child Size from Birth to Age 14 Years *P. D. Sampson, F. L. Bookstein, H. M. Barr, and A. P. Streissguth*
1429 Infertility in Women and Moderate Alcohol Use *F. Grodstein, M. B. Goldman, and D. W. Cramer*
1433 Trends in the US Prevalence of Drug-Using Parturient Women and Drug-Affected Newborns, 1979 through 1990 *M. Dicker and E. A. Leighton*
1439 Black–White Differences in Serum Cotinine Levels among Pregnant Women and Subsequent Effects on Infant Birthweight *P. B. English, B. Eskenazi, and R. E. Christianson*
1444 Birthweight Differentials among Asian Americans *G. K. Singh and S. M. Yu*
1450 The Relationship of Prenatal Care and Pregnancy Complications to Birthweight in Winnipeg, Canada *C. A. Mustard and N. P. Roos*
1458 Is Breast-Feeding in Infancy Associated with Adult Longevity? *D. L. Wingard, M. H. Criqui, S. L. Edelstein, J. Tucker, C. Tomlinson-Keasey, J. E. Schwartz, and H. S. Friedman*
1463 Harsh Social Conditions and Perinatal Survival: An Age-Period-Cohort Analysis of the World War II Occupation of Norway *A. J. Wilcox, R. Skjaerven, and L. M. Irgens*

(continued on page 1361)

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

American Public Health Association

September 1994, Vol. 84, No. 9 ISSN: 0090-0036 Established 1911

1468 State Family Planning and Abortion Expenditures: Their Effect on Public Health *K. J. Meier and D. R. McFarlane*

1473 The Costs and Financing of Perinatal Care in the United States *S. H. Long, M. S. Marquis, and E. R. Harrison*

Public Health Briefs

VARIOUS TOPICS

1479 Targeting Cancer Control: The State Cancer Control Map and Data Program *F. A. Pommerenke, R. W. Miller, S. Srivastava, and S. P. Ackermann*

1483 Elevated Blood Lead Levels among Construction Workers in the Massachusetts Occupational Lead Registry *R. Rabin, D. R. Brooks, and L. K. Davis*

PERINATAL HEALTH

1486 The Adequacy of Prenatal Care Utilization Index: Its US Distribution and Association with Low Birthweight *M. Kotelchuck*

1489 The Effect of Paternal Smoking on the Birthweight of Newborns Whose Mothers Did Not Smoke *F. D. Martinez, A. L. Wright, L. M. Taussig, and the Group Health Medical Associates*

1492 Intrauterine Cocaine, Lead, and Nicotine Exposure and Fetal Growth *D. R. Neuspiel, M. Markowitz, and E. Drucker*

1495 Selected Antepartum Medical Complications and Very-Low-Birthweight Infants among Black and White Women *M. DeBaun, D. Rowley, M. Province, J. W. Stockbauer, and F. S. Cole*

1497 Infant Mortality and Related Risk Factors among Asian Americans *H. W. Morrow, G. F. Chávez, P. P. Giannoni, and R. S. Shah*

1501 The Perinatal and Infant Health Status of Native Hawaiians *E. C. Kieffer, J. M. Mor, and G. R. Alexander*

1505 Pregnancy Risk Factors and Birth Outcomes in Washington State: A Comparison of Ethiopian-Born and US-Born Women *H. Wasse, V. L. Holt, and J. R. Daling*

1508 The Risk of Pregnancy-Induced Hypertension: Black and White Differences in a Military Population *D. E. Irwin, D. A. Savitz, I. Hertz-Picciotto, and K. A. St. André*

1511 Nonclient Factors in the Reporting of Prenatal Psychosocial Risk Assessments *D. S. Wilkinson, C. C. Korenbrot, and E. Fuentes-Afflick*

Health Law and Ethics

1515 Patients' Rights after Health Care Reform: Who Decides What Is Medically Necessary? *W. K. Mariner*

Other Journal Departments

1362 APHA Masthead

1365 Advertisers' Index

1366 What *AJPH* Authors Should Know

1369 Journal Masthead

1521 Letters to the Editor

1527 Book Corner

1531 Annual Meeting Placement Service Forms

1541 Annual Meeting Registration Form and Housing Information

1546 Job Opportunities

American Journal of Public Health

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 15th St, NW, Washington, DC 20005; (202) 789-5600; TDD (202) 789-5673. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*
Helen Rodriguez-Trias, MD, *Immediate Past President*
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Leonard M. Schuman, MD, *Vice-President (USA)*
John E. F. Hastings, MD, MPH, *Vice-President (Canada)*
Oscar Feo, MD, MSc, *Vice-President (Latin America)*
Jay H. Glasser, PhD, *Treasurer*
Allan G. Rosenfield, MD, *Chair, Executive Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*
Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*
Carol Easley Allan, PhD, RN (1997)
Caswell A. Evans, Jr, DDS, MPH, *President-Elect*
Patricia P. Evans, MPH (1994)
Eugene Feingold, PhD, JD, *President*
Jay H. Glasser, PhD, *Treasurer*
Alan R. Hinman, MD, MPH (1995)
Margo D. Kerrigan, MPH (1995)
Joanne Lukomnik, MD, MPH (1996) *Vice-Chair*
Steven Uranga McKane, DMD, MPH (1995)
Carlos W. Molina, EdD (1997)
Linda Rae Murray, MD, MPH (1996)
Hila Richardson, DrPH, MPH (1996)
Helen Rodriguez-Trias, MD, *Immediate Past President*
Reed V. Tuckson, MD (1994)
Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*
Frances Kissling, *Chair, Action Board*
George E. Pickett, MD, MPH, *Speaker, Program Development Board*
Shirley F. Randolph, MSPH, *Speaker, Governing Council*

The *American Journal of Public Health* (*AJPH*), published monthly, is the official Journal of the American Public Health Association, Inc, 1015 15th St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005. *AJPH* allotted $40; *The Nation's Health* allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters. Printed in the United States.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on the first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, *AJPH* Reprints, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the Journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The *American Journal of Public Health* is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, ETOH, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

Community Education for Cardiovascular Disease Prevention: Risk Factor Changes in the Minnesota Heart Health Program

Russell V. Luepker, MD, David M. Murray, PhD, David R. Jacobs, Jr, PhD, Maurice B. Mittelmark, PhD, Neil Bracht, MSW, MPH, Ray Carlaw, DrPH, Richard Crow, MD, Pat Elmer, PhD, John Finnegan, PhD, Aaron R. Folsom, MD, Richard Grimm, MD, PhD, Peter J. Hannan, MStat, Robert Jeffrey, PhD, Harry Lando, PhD, Paul McGovern, PhD, Rebecca Mullis, PhD, Cheryl L. Perry, PhD, Terry Pechacek, PhD, Phyllis Pirie, PhD, J. Michael Sprafka, PhD, Rita Weisbrod, PhD, and Henry Blackburn, MD

ABSTRACT

Objectives. The Minnesota Heart Health Program is a 13-year research and demonstration project to reduce morbidity and mortality from coronary heart disease in whole communities.

Methods. Three pairs of communities were matched on size and type; each pair had one education site and one comparison site. After baseline surveys, a 5- to 6-year program of mass media, community organization, and direct education for risk reduction was begun in the education communities, whereas surveys continued in all sites.

Results. Many intervention components proved effective in targeted groups. However, against a background of strong secular trends of increasing health promotion and declining risk factors, the overall program effects were modest in size and duration and generally within chance levels.

Conclusions. These findings suggest that even such an intense program may not be able to generate enough additional exposure to risk reduction messages and activities in a large enough fraction of the population to accelerate the remarkably favorable secular trends in health promotion activities and in most coronary heart disease risk factors present in the study communities. (*Am J Public Health.* 1994;84:1383–1393)

Introduction

Cardiovascular diseases, principally coronary heart disease and stroke, remain the leading causes of death and disability in industrialized countries. They are preventable and reversible through the modification of risk factors, including elevated blood lipids, elevated blood pressure, cigarette smoking, and sedentary life-style.[1–4] Entire populations are at increased risk because of mass elevated risk factors in which individual susceptibility is enhanced by culture, economic factors, and the environment. Population risk should be amenable to change through communitywide strategies,[5] and several community intervention research and demonstration projects are under way or completed in the United States and Europe.[6–12]

The largest of these studies in the United States is the Minnesota Heart Health Program. Initiated in 1980, the Minnesota Heart Health Program involves approximately 400 000 persons in six communities in the Upper Midwest. It was hypothesized that a 5- to 6-year intervention program would (1) improve health behaviors; (2) lower population levels of blood cholesterol by 7 mg/dL, blood pressure by 2 mm Hg, and cigarette smoking by 3% and increase population physical activity levels by 50 kcal/day; and (3) that these changes would subsequently reduce cardiovascular disease morbidity and mortality by 15%. Risk factor levels and associated behaviors were measured before the start of the intervention program and annually for 6 or 7 years of follow-up in all six communities; morbidity and mortality data are being collected annually from the six communities and surrounding states. The purpose of this paper is to present the main risk factor results of the study.

Methods

Study Design

The design of the Minnesota Heart Health Program is described elsewhere[6] and summarized here (Figure 1). Three pairs of communities were selected; each pair had one education site and one

Russell V. Luepker, David M. Murray, David R. Jacobs Jr, Ray Carlaw, Richard Crow, Pat Elmer, John Finnegan, Aaron R. Folsom, Peter J. Hannan, Robert Jeffrey, Harry Lando, Paul McGovern, Cheryl L. Perry, Phyllis Pirie, J. Michael Sprafka, and Henry Blackburn are with the Division of Epidemiology, School of Public Health, University of Minnesota, Minneapolis, Minn. Maurice B. Mittelmark is with the Bowman Gray School of Medicine, Wake Forest University, Winston Salem, NC. Neil Bracht is with the School of Social Work, University of Minnesota, Minneapolis. Richard Grimm is with the School of Medicine, University of Minnesota, Minneapolis. Rebecca Mullis is with the Division of Nutrition, Centers for Disease Control and Prevention, Atlanta, Ga. Terry Pechacek is with the Department of Social and Preventive Medicine, State University of New York at Buffalo, Buffalo, NY. Rita Weisbrod is with the Department of Sociology, Augsburg College, Minneapolis, Minn.

Requests for reprints should be sent to Russell V. Luepker, MD, Division of Epidemiology, School of Public Health, University of Minnesota, 1300 S Second St, Suite 300, Minneapolis, MN 55455-1015.

This paper was accepted April 6, 1994.

Editor's Note. See related editorial by Winkleby (p 1369) in this issue.

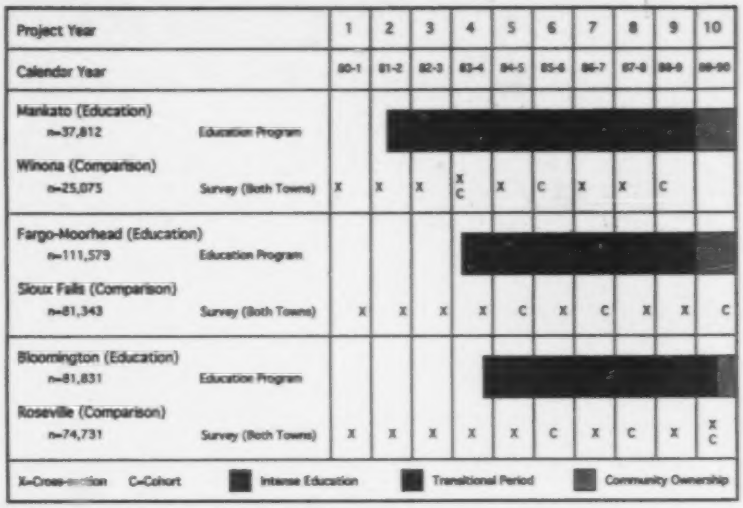
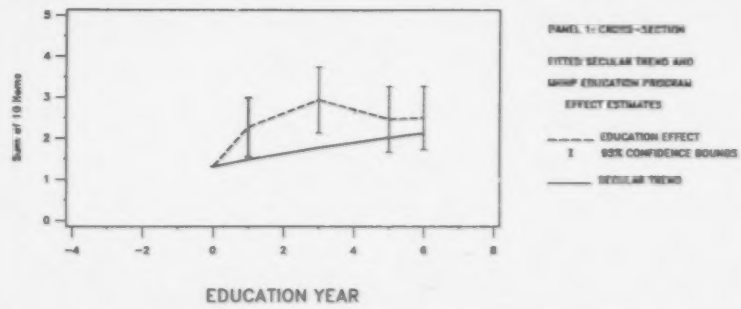


FIGURE 1—Minnesota Heart Health Program study design. Reprinted from Murray et al.[80] with permission. Copyright © 1994 *American Journal of Epidemiology*.



Note. Adjusted for age, gender, and education. MHHP = Minnesota Heart Health Program.

FIGURE 2—Exposure score: results of the MHHP education intervention program.

comparison site matched on size (25 000–110 000), community type (small, medium, urban), and distance from the Minneapolis–St. Paul metropolitan area (<250 miles). Assignment of communities to intervention within pairs was nonrandom and completed before collection of the data. After a 16-month baseline period, a 5- to 6-year intervention program was introduced in November 1981 in Mankato, Minnesota, a rural community in southern Minnesota. The program was introduced 22 and 28 months later, respectively, in Fargo–Moorhead, an urban area consisting of two neighboring communities along the North Dakota–Minnesota border, and in Bloomington, Minnesota, a large Minneapolis–St. Paul suburb. The staggered entry allowed for gradual development of the intervention program and strengthened the design through replication; it also provided two, three, and four baseline surveys in the first, second, and third pairs, respectively, to improve the precision of the preintervention time trends estimated from the data.

Survey Methods

Cross-sectional surveys. Cross-sectional surveys of 300–500 randomly selected adults were conducted periodically in each community, with a two-stage cluster sampling design.[13] Initially, census blocks were randomly selected from each city, with the probability of selection proportional to the expected number of households. Geographically adjacent groups of five households were randomly selected from within those blocks. Within households, a single age-eligible adult was selected at random. This procedure was repeated for the second survey sample; subsequent samples were drawn from the census blocks selected in the first 2 years. Matched communities were surveyed in adjacent 2-month periods.

After selection, initial household contact was by letter, followed by direct interview to (1) enumerate the household; (2) randomly select one age-eligible subject; (3) collect data on sociodemographic characteristics, medical history, health attitudes and beliefs, and smoking history; and (4) extend an invitation to visit a survey center for risk factor measurements. Those who spoke no English or who were judged mentally incompetent to participate were ineligible (<5%), and some participants were examined at home (<5%).

Cohort surveys. Periodic remeasurements were made in a cohort comprised of participants selected at random from all of the preintervention cross-sectional surveys. To minimize the effect of repeated testing, remeasurements were structured so that approximately half the cohort was selected at random for recontact after 2 years of intervention, whereas the other half was recontacted after 4 years. The entire cohort was recontacted after 6–7 years of intervention. As expected, there were no differences in baseline risk factors between the two halves of the cohort.

Measurements

Detailed methods for risk factor measurements are reported elsewhere[14] and are only summarized here.

Blood pressure was measured in a seated position with a random-zero sphygmomanometer (Hawksley–Gelman, London, England) and an appropriate-sized cuff after a 5-minute rest. Two measurements of systolic and fifth-phase diastolic pressure were made 1 minute apart and averaged for the analysis.

Blood was obtained according to the methods described in the Lipid Research Clinic's protocol.[15] Serum total cholesterol was measured in the Minnesota Lipid Research Clinic's laboratory by an Autoanalyzer II (Technicon Instrument

TABLE 1—Pooled Education and Comparison City-Year Means from the Cross-Sectional Surveys, Adjusted within Strata and Standardized across Strata, before Modeling of Secular Trends and Minnesota Heart Health Program Intervention Effects

		Education Year[a]							
		−3[b]	−2[b]	−1	0	1	3	5	6
Exposure score (range, 1–10)	Education				1.3	2.3	2.9	2.4	2.5
	Comparison				1.2	1.6	1.8	1.9	2.1
Blood cholesterol, mg/dL	Education	213.5	207.2	206.5	207.3	199.8	199.0	204.1	203.8
	Comparison	212.3	208.2	209.3	208.3	208.8	199.4	199.9	203.8
Smoking, males (% smokers)	Education	38.7	45.2	34.5	37.9	34.0	33.7	30.4	29.4
	Comparison	34.5	43.8	36.3	39.5	32.3	31.6	27.2	31.1
Smoking, females (% smokers)	Education	25.3	31.6	33.5	28.0	30.5	23.8	22.8	22.8
	Comparison	22.1	31.0	29.3	27.8	26.7	23.4	28.3	26.0
Systolic blood pressure, mm Hg	Education	120.8	120.7	121.5	121.7	121.6	118.2	119.4	118.6
	Comparison	124.5	122.6	124.3	125.7	124.4	121.7	120.5	121.6
Diastolic blood pressure, mm Hg	Education	74.6	73.8	75.2	76.2	75.4	71.6	73.4	73.3
	Comparison	76.5	74.0	76.6	76.8	75.1	74.5	74.1	74.1
Body mass index, kg/m^2	Education	25.3	25.6	25.8	25.6	25.7	26.0	26.3	26.0
	Comparison	25.5	25.7	25.8	25.8	26.3	26.2	26.4	26.5
Physical activity (% active)	Education	50.9	49.5	45.5	48.4	54.2	54.3	55.2	57.1
	Comparison	52.1	49.0	50.3	47.3	48.4	51.3	53.9	52.8
Coronary heart disease risk (deaths/1000 persons)	Education	28.7	26.9	26.5	25.5	26.3	22.6	23.8	24.4
	Comparison	32.2	35.3	33.7	33.6	34.3	29.7	28.1	31.2

[a]Negative values refer to observations made before the intervention program; a value of 0 refers to observations made immediately before the intervention program; positive values refer to observations made after the intervention program began.
[b]Data were not collected from Mankato, Winona in E-3 or E-2, nor in Fargo/Moorhead, Sioux Falls in E-3; pooled means for E-3 and E-2 were calculated after imputing values for the missing components based on the differences in the city levels as estimated in the regressions.

Corporation, Tarrytown, NY).[16] Quality control of Lipid Research Clinic laboratories was maintained by external standardization with the Centers for the Disease Control (CDC),[16] and the values were adjusted for daily variability against standard CDC pools.[17]

Participants were classified as current smokers if they had ever smoked at least 100 cigarettes in their lifetime and smoked at present. Serum thiocyanate was also measured.[18]

Height and weight were measured in stocking feet and light clothing. Height was measured to the nearest centimeter with a fixed metal rule. Weight was measured to the nearest half pound with a balance beam scale. Body mass index was computed as kilograms per square meter.

Leisure time physical activity was assessed as the percentage of participants who answered "yes" to the question, "Are you regularly active in your leisure time?"

A risk reduction exposure score, which was constructed from 10 questions about recent participation in activities such as smoking cessation classes, cholesterol screening, and so forth, was added when intervention began. The score ranged from 0 to 10, and all items were designed to be equally applicable in all communities.

Other variables were employed as stratification factors (age, gender, educational attainment) or as covariates (specific to each outcome and identified in Figures 2–10).

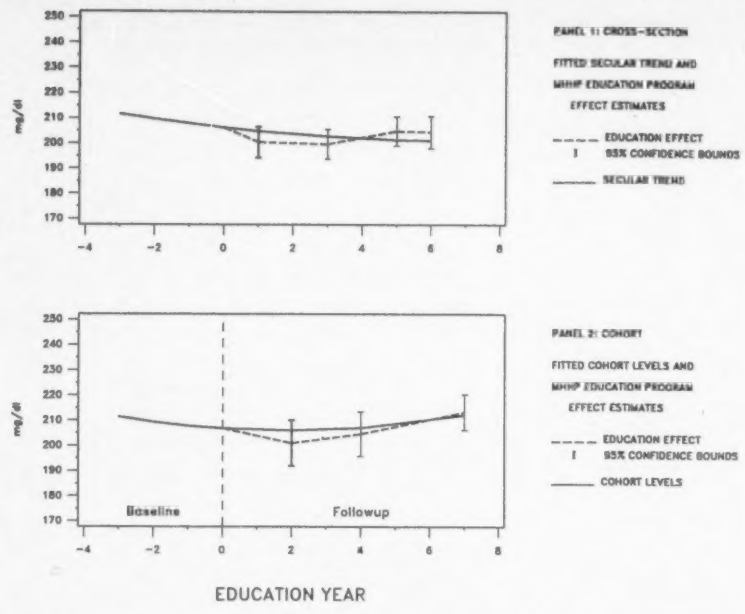
Scores for the risk per 1000 persons of death from coronary heart disease in 10 years were computed according to the method of Truett et al.[19] with the coefficients provided by Leaverton et al.[20]; separate functions were used for men and women.* The variables included were age in years, systolic blood pressure, total cholesterol, and cigarette smoking.

Intervention Methods

The Minnesota Heart Health Program intervention advocated hypertension prevention and control, healthy eating patterns for lower blood cholesterol and blood pressure, nonsmoking, and regular physical activity. It operated at the individual, group, and community levels and embraced a wide range of strategies and theories, including social learning theory,[21] persuasive communications theory,[22,23] and models for involvement of community leaders and institutions.[24] The program alerted people to health issues, informed them of effective behavioral alternatives for health promotion, provided incentives for new behaviors, and provided reinforcements to maintain new behaviors.

Community analysis and organization methods were used to engage community leaders and organizations as active participants in the intervention programs; this effort resulted in their active involvement, gradual environmental change to support risk reduction, and community planning for program continuation.[25–27] Mass media were used to increase individual exposure to the Minnesota Heart Health Program risk factor messages, establish awareness of the program, and increase the salience of the program

*Risk = $1000 * (1/1 + e^{-c})$.
For men: $c = -9.42705 + 0.0644(\text{age}) + 0.0131(\text{SBP}) + 0.0046(\text{CHOL}) + 0\ 7809(\text{SMK})$.
For women: $c = -15.1298 + 0.1059(\text{age}) + 0.0223(\text{SBP}) + 0.0097(\text{CHOL}) + 0.8823(\text{SMK})$.



Note. Adjusted for age, gender, education, and seasonal trend. MHHP = Minnesota Heart Health Program.

FIGURE 3—Total cholesterol: results of the MHHP education intervention program.

messages.[28–32] Health professionals in the education communities were involved through their local organizations and preventive practice advisory committees, and they served as role models and opinion leaders.[33,34] Systematic risk factor screening and education were conducted during the first 3 years of the intervention program; over 60% of all adult residents received on-site measurement, education, and counseling.[35] The adult education component made available personal, intensive, and multiple contact programs to reduce cardiovascular risk; this strategy focused on self-management and included changes in existing behaviors, in the meaning of those behaviors, and in the environmental cues that supported those behaviors.[36–44] Direct education programs for school-age children discouraged health-compromising behaviors and promoted health-enhancing behaviors in youth and their parents.[45–47]

The program included a high-intensity campaign via the mass media, recruited virtually all primary care physicians and many other health professionals to training programs, recruited more than 60% of all adults age 25–74 to the screening and education programs, recruited more than 30% of all adults to

TABLE 2—Pooled Education and Comparison City-Year Means from the Cohort Surveys, Adjusted within Strata and Standardized across Strata, before Modeling of Trends or Minnesota Heart Health Program Intervention Effects

		Education Year[a]						
		−3[b]	−2[b]	−1	0	2	4	7
Blood cholesterol, mg/dL	Education	216.3	208.7	206.3	206.8	201.3	202.5	213.0
	Comparison	212.3	205.7	209.4	209.7	207.6	207.0	213.0
Smoking, males (% smokers)	Education	30.8	39.0	27.0	32.2	27.5	26.0	21.4
	Comparison	26.6	36.2	27.7	34.4	24.7	26.7	19.6
Smoking, females (% smokers)	Education	23.8	32.9	30.8	26.7	24.7	22.9	19.6
	Comparison	23.2	28.1	30.5	27.9	25.0	24.1	19.6
Systolic blood pressure, mm Hg	Education	121.0	120.4	121.3	122.3	121.2	120.4	121.3
	Comparison	123.1	122.2	123.9	126.0	124.0	123.4	123.6
Diastolic blood pressure, mm Hg	Education	74.7	73.1	74.6	76.2	74.8	75.0	75.8
	Comparison	75.5	73.9	76.0	76.8	76.3	74.9	75.8
Body mass index, kg/m^2	Education	25.7	25.7	25.9	25.8	26.0	26.5	26.8
	Comparison	25.4	25.8	25.6	25.9	26.2	26.5	26.9
Physical activity (% active)	Education	51.8	49.3	45.0	49.4	56.6	57.4	63.6
	Comparison	55.0	48.3	51.5	49.1	58.4	58.2	58.6
Coronary heart disease risk (deaths/1000 persons)	Education	27.0	25.4	24.4	24.7	30.7	32.3	41.0
	Comparison	31.2	32.5	31.6	31.0	38.2	43.4	50.4

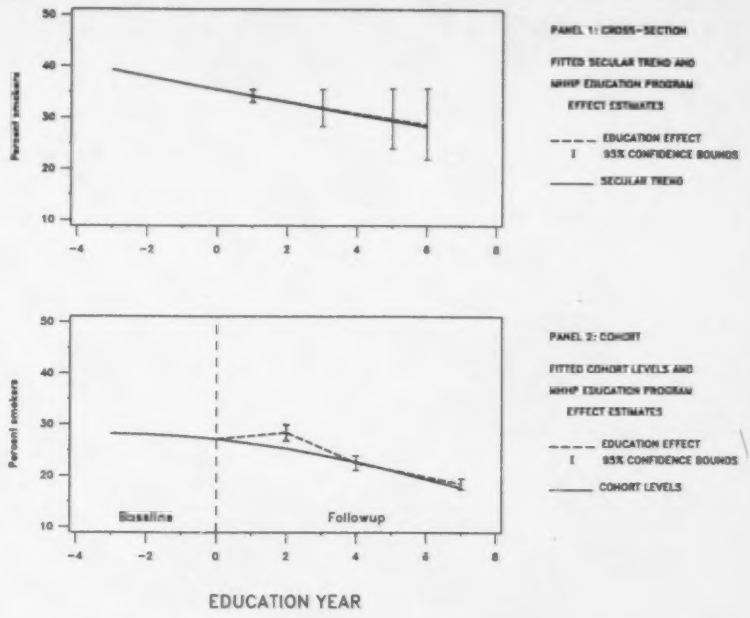
[a]Negative values refer to observations made before the intervention program; a value of 0 refers to observations made immediately before the intervention program; positive values refer to observations made after the intervention program began.
[b]Data were not collected from Mankato, Winona in E-3 or E-2, nor in Fargo/Moorhead, Sioux Falls in E-3; pooled means for E-3 and E-2 were calculated after imputing values for the missing components based on the differences in city levels as estimated in the regressions.

face-to-face intervention programs, involved organizations in environmental change programs, and involved most of the resident youth in school-based health promotion activities.

Analysis Methods

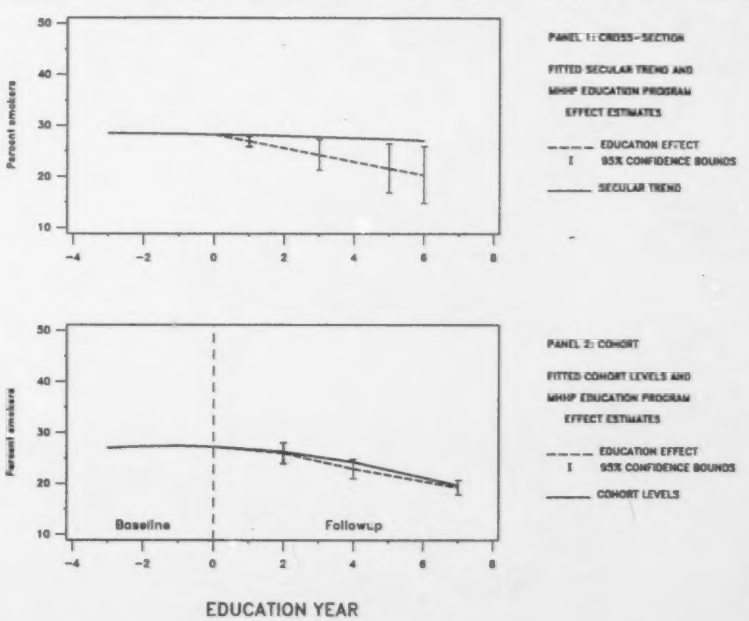
By design, the community was the unit of assignment, whereas the individual was the unit of observation. Persons within clusters such as communities tend to be more like one another than they are like persons in other communities,[13] and this within-cluster correlation adds an additional component to the variability of the treatment group means over that attributable either to the individual participants or to the treatments.[48] Unless this extra variation is accounted for in analysis, the *P* values for the treatment effects will be artificially deflated.[49] We accounted for the extra variation through a two-stage analysis that approximated a stratified hierarchical analysis of covariance; the nested communities were treated as random effects; condition, time, and stratification factors were treated as fixed effects; and all tests of statistical significance were made at the community level.[50,51]

In the first stage, least-squares adjusted means were generated for each city in each survey year after stratifying simultaneously by gender, educational attainment, and age; adjustments were made for confounding variables specific to each dependent variable by standardizing all subjects to the population average within strata for each covariate. The second stage employed these adjusted strata-specific city-year means as the unit of analysis in a series of regressions to evaluate the main and strata-specific effects of the intervention program as a time-by-treatment interaction. Recognizing that planned contrasts can provide tests that are more interpretable and potentially more powerful, we tested two specific patterns of program effect. For both patterns, a quadratic secular trend was modeled by using the baseline intervention city-year means together with all of the comparison city-year means. The intervention program effect was then modeled first as a series of year-specific departures from that secular trend and second as a linear departure from that trend. The analysis of the cohort data followed the same plan; modifications were made to accommodate time-varying covariates and the fact that different members of the cohort participated in each of the four cohort surveys.



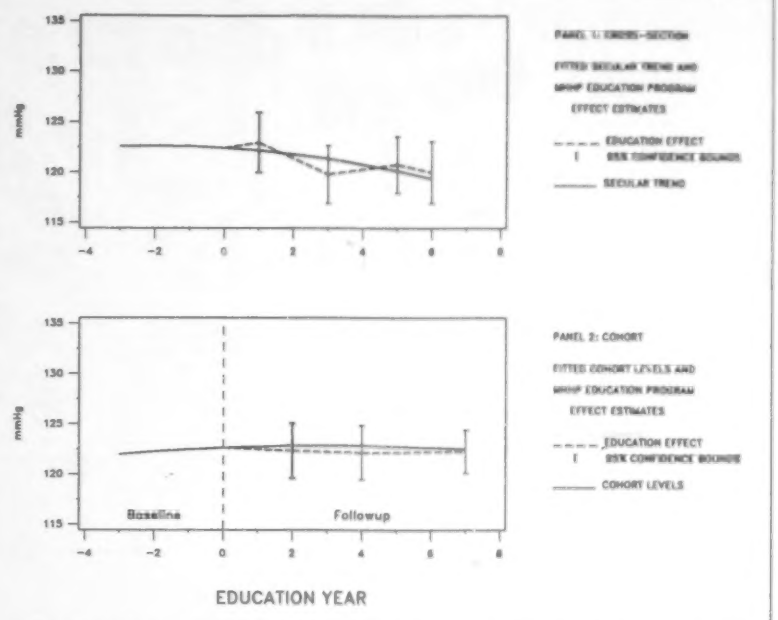
Note. Adjusted for age, education, and marital status. MHHP = Minnesota Heart Health Program.

FIGURE 4—Cigarette smoking in males: results of the MHHP education intervention program.



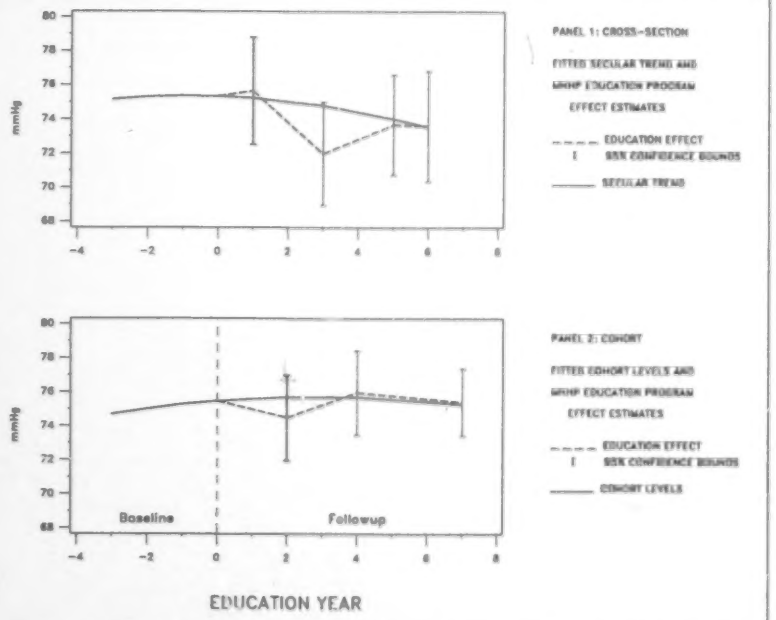
Note. Adjusted for age, education, and marital status. MHHP = Minnesota Heart Health Program.

FIGURE 5—Cigarette smoking in females: results of the MHHP education intervention program.



Note. Adjusted for age, gender, education, seasonal trend, cuff size, observer, and location of blood pressure measurement. MHHP = Minnesota Heart Health Program.

FIGURE 6—Systolic blood pressure: results of the MHHP education intervention program.



Note. Adjusted for age, gender, education, seasonal trend, cuff size, observer, and location of blood pressure measurement. MHHP = Minnesota Heart Health Program.

FIGURE 7—Diastolic blood pressure: results of the MHHP education intervention program.

Results

Survey Participation

A total of 20 184 eligible adults age 25–74 years completed the home interview, and 18 062 completed the survey center protocol. Participation rates were stable across all years and communities: enumeration averaged 96.1% of the selected households, home interviews averaged 87.9% of the selected adults, and the total survey response averaged 78.7%.

Of the participants from the baseline cross-sectional surveys, 7097 were randomly selected for the cohort surveys; the refusal rate was 14.9% and the population was relatively stable, with 67.1% of the original group living in the town and participating at the end of the study. The remainder moved or died. Those lost to follow-up were compared with those retained. The ones lost to follow-up had higher smoking rates and lower levels of blood cholesterol, but there was no evidence of differential attrition between the education and comparison sites. The only differences between the education and the comparison communities on loss to follow-up were in body mass index at the 7-year follow-up survey; those lost from the education communities were slightly leaner than those lost from the comparison communities (body mass index of 25.4 vs 25.9).

Exposure to Risk Reduction Activities

Figure 2 summarizes exposure to risk reduction activities, and Table 1 presents the year-specific means for the exposure summary score; only cross-sectional data are presented because the exposure measures were not included in all cohort surveys. The solid line represents the secular trend estimated for exposure based on all nonintervention stratum-specific city-year means and indicates that exposure to these risk reduction activities increased over time in the comparison communities. The dashed line illustrates the Minnesota Heart Health Program effects at 1, 3, 5, and 6 years of intervention. Exposure was significantly higher in the education communities compared with the comparison communities after 1 and 3 years, but not after 5 or 6 years.

Blood Cholesterol Level

Figure 3 and Tables 1 and 2 summarize the findings for blood cholesterol. In the cross-sectional data, the secular trend showed a steady decline in the comparison communities (about −1.12 mg/dL per

year). The mean blood cholesterol was lower in the education communities during the first 3 years of the intervention program and higher in the final 2 years. None of these differences were statistically significant. The cohort findings were similar.

Smoking Prevalence

Figures 4 and 5 and Tables 1 and 2 summarize the findings for cigarette smoking. The response to the intervention program differed by gender, so the figures and tables are gender specific. Among men, there was a downward trend in the comparison communities, both in the cross-section and in the cohort (about −1.5% per year), and there was no added intervention effect. Among women, the secular trend was small in the cross-section, and there was a statistically significant treatment effect (about −1.4% per year) that remained after we corrected the smoking self-report data for underreporting by using serum thiocyanate levels (data not shown). In the cohort, there was a marked decline in smoking prevalence among women and only limited evidence of an intervention effect.

Blood Pressure

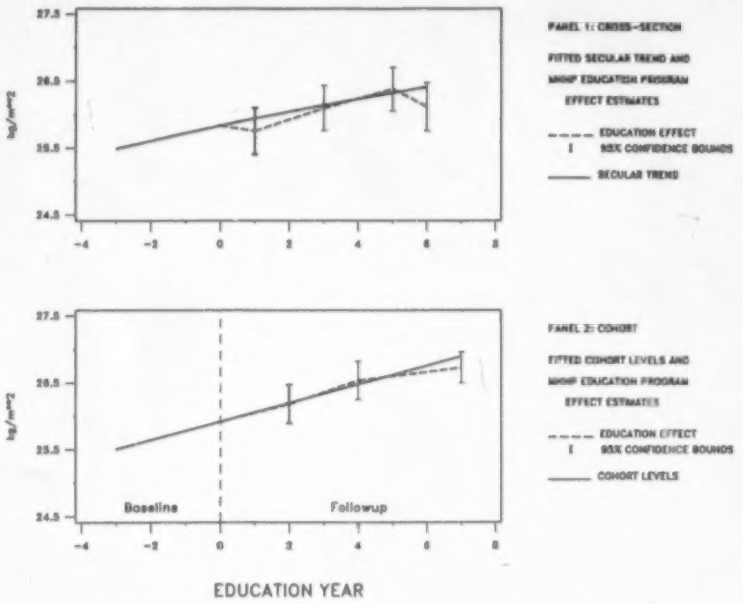
Figures 6 and 7 and Tables 1 and 2 summarize the findings for blood pressure. In the comparison communities, there was evidence of a declining secular trend in the cross-section (averaging −0.4 mm Hg for systolic blood pressure and −0.2 mm Hg for diastolic blood pressure per year) and a flat or slightly rising trend in the cohort, as would be expected with aging. In the education communities, both measures had an additional modest decline after 3 years of intervention in the cross-section and after 2 or 4 years in the cohort; however, none of these changes were statistically significant.

Body Mass Index

Figure 8 and Tables 1 and 2 summarize the findings for body mass index. Both the secular and cohort trends were positive (about +0.1 kg/m^2 per year), and there was no evidence of an intervention effect.

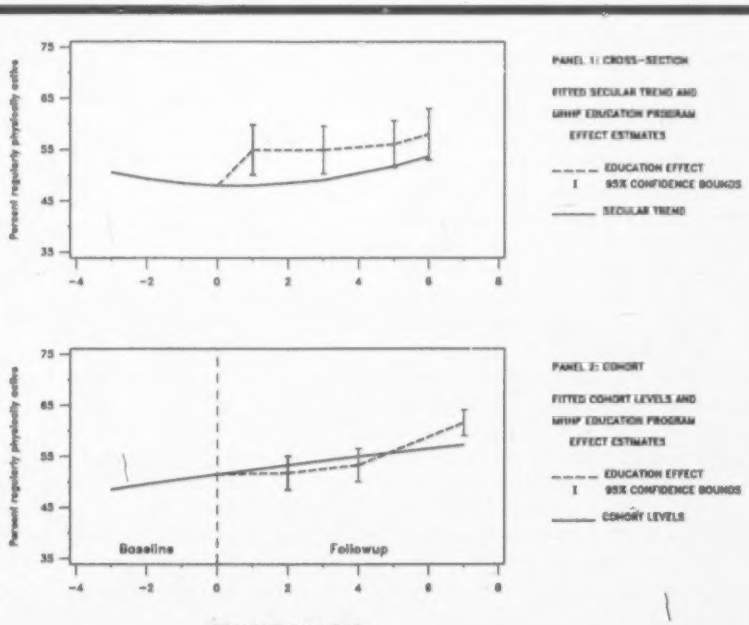
Regular Physical Activity

Figure 9 and Tables 1 and 2 summarize the findings for regular physical activity. The secular trend declined initially but then rose over the last 4 years of the study. The Minnesota Heart Health



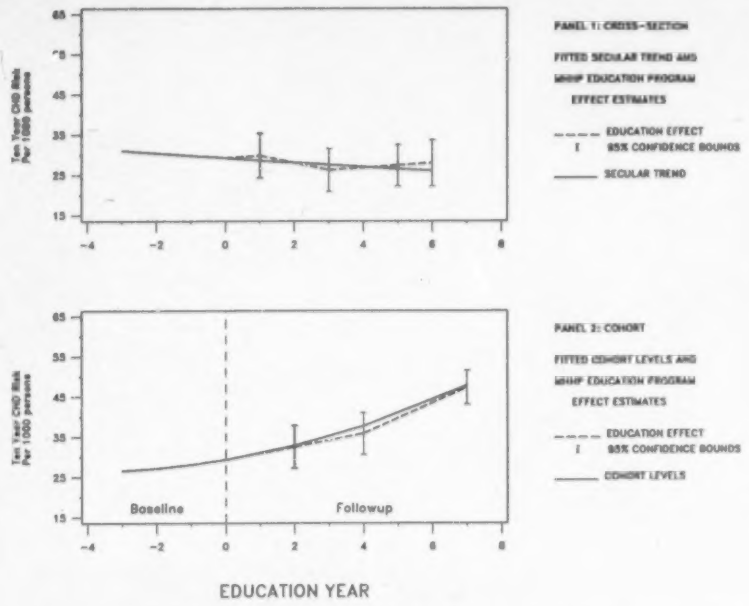
Note. Adjusted for age, gender, education, marital status, and seasonal trend. MHHP = Minnesota Heart Health Program.

FIGURE 8—Body mass index: results of the MHHP education intervention program.



Note. Adjusted for age, gender, and education. MHHP = Minnesota Heart Health Program.

FIGURE 9—Regular physical activity: results of the MHHP education intervention program.



Note. Risk scores were computed by applying the coefficients of Leaverton et al.[20] to the predicted risk factor levels for each participant obtained from the individual risk factor analyses. CHD = coronary heart disease; MHHP = Minnesota Heart Health Program.

FIGURE 10—Coronary heart disease risk score: results of the MHHP education intervention program.

intervention program accelerated this secular trend significantly, especially during the first 3 years of the intervention program. The cohort data suggest a steady increase in self-reported regular physical activity in all communities over the course of follow-up; the education communities exceeded the comparison communities only at the last follow-up survey. We also analyzed the more extensive physical activity data based on the Minnesota Leisure Time Physical Activity Questionnaire, which was answered by a random half of the Minnesota Heart Health Program survey participants. Unlike the findings based on the single question concerning regular participation in leisure time physical activity, results of the longer questionnaire showed a small increase in kilocalories per day expended in leisure time physical activity in the early years of the education program, but a small decrease in later years (data not shown). Time spent in heavier-intensity activities appeared to decrease slightly in the education communities by the end of the program. Resting pulse rate trends (data not shown) reflected the self-reported changes, with lower pulse rates in the education communities associated with higher reported physical activity levels.

Coronary Heart Disease Risk

Figure 10 and Tables 1 and 2 summarize the risk score findings. Risk of death from coronary heart disease declined over time in the cross-section in both the education and the comparison communities, but there was little evidence of an intervention effect. The cohort data reflected increasing risk as the cohort aged, but again, there was little evidence of an intervention effect.

Discussion

The differences observed here between the education and the comparison communities with respect to population-wide risk factor changes were less than postulated. There are several potential explanations for these findings. First, it is apparent that there were strong and favorable secular trends of both increasing health promotion activities and declining risk factors for coronary heart disease in all study communities. Second, it is apparent that the net improvements in health promotion activities and individual risk factors that can be attributed to the Minnesota Heart Health Program intervention were modest, generally of limited duration, and usually within chance levels. In light of the results from previous studies that have supported the efficacy of many of the Minnesota Heart Health Program intervention components in targeted populations, we cannot conclude that those components are ineffective. Instead, we conclude that the Minnesota Heart Health Program intervention was unable to generate enough additional exposure to those risk reduction activities in a large enough proportion of the population to exceed the remarkably favorable secular trends that were ongoing in the study communities. In the next several paragraphs, we review the evidence that led us to this conclusion, consider how these findings compare with the results from similar studies, and consider the implications for public health practice and policy.

Consider first the data on the secular trends for exposure to health promotion activities related to coronary heart disease (Figure 2). Those results suggest that exposure grew steadily in both the comparison and education communities. Examination of the individual items in the exposure score revealed that the net increase in the education communities was attributable largely to higher rates of exposure to cholesterol screening, health projects or surveys, and restaurant menu labeling programs. Other exposure measures showed little difference between the education and the comparison communities. These results suggest that in spite of the Minnesota Heart Health Program's intense intervention program, the increase in such activity over the growing levels observed in the comparison communities was surprisingly limited. This interpretation must be tempered by the limitations inherent in the exposure measures themselves, which assessed whether or not respondents had been exposed to a particular type of risk reduction message or activity and did not discriminate among exposures on the basis of their quality, intensity, or duration. Even so, the exposure data suggest that the Minnesota Heart Health Program may not have added a great deal to the level of risk reduction activity that would have been expected without the program. It may be that the Minnesota Heart Health Program represented too small a fraction of the total exposure to messages related to coronary heart disease to make much difference, or that the Minnesota Heart Health Program supplanted activities that

might otherwise have developed on their own. Either way, the net gain in exposure was modest.

Consistent with the favorable secular trend in exposure to risk reduction messages and activities, the cross-sectional results indicate that there were beneficial and often strong trends for each risk factor, with only two exceptions (smoking in women and body mass index). For example, smoking prevalence declined in men by 11.3%, systolic blood pressure declined by 6.3 mm Hg, and total blood cholesterol declined by 11 mg/dL. The major surprise with regard to these secular trends was their magnitude, because only small changes had been expected in these risk factors when the study was designed in 1978. Instead, the improvements in the secular trends were often much greater than the intervention effects hypothesized for the Minnesota Heart Health Program (e.g., hypothesized effects were for a 3% decline in smoking prevalence, a 7 mg/dL decline in total blood cholesterol, and a 2 mm Hg decline in systolic blood pressure). We note as well that the secular trends were greater than those previously observed nationally in the National Health and Nutrition Examination Survey[52] or regionally in the Minnesota Heart Survey.[53]

The cross-sectional results suggest intervention effects in self-reported smoking prevalence among women and in self-reported physical activity in both men and women; however, the cohort data did not support intervention effects in any of the risk factors. In spite of an extensive effort to explain the difference in smoking outcomes for women between the cross-section and the cohort, no explanation has been found. Because smoking in women was the only risk factor that showed a flat trend over time in the comparison communities, this finding may indicate that the Minnesota Heart Health Program intervention was able to induce a favorable slope in the absence of a favorable secular trend. The increased effect observed toward the end of the study is also consistent with the increased delivery of antismoking messages and activities late in the intervention program. The follow-up analyses for physical activity also suggested that the apparent increase in the proportion of the population who engaged in physical activity in their leisure time was the result of an increase in the proportion who engaged in light activity, not an increase in the proportion who engaged in moderate or vigorous activity. This may explain the absence of a Minnesota Heart Health Program effect on body mass index in spite of the apparent effect on physical activity. Taken together, the risk factor results are consistent with the exposure data and suggest modest variation in coronary heart disease risk factors around their secular trends; the data provide little evidence of any broad or significant acceleration of those secular trends that can be attributed with confidence to the intervention program.

Because data on morbidity and mortality are not yet available, we must look to the coronary heart disease risk score for preliminary evidence of the net effect of the intervention program on the population risk for coronary heart disease (Figure 10). The secular trend in the cross-sectional data is consistent with the trends seen in the exposure and individual risk factor data: the estimated 10-year risk improved by 5.0 predicted deaths per 1000 persons from coronary heart disease over the 9 years of the study. The intervention effect is also consistent with the previous results. As with the risk factors, the improvement in estimated risk due to the secular trend was also larger than the improvement that had been hypothesized for the Minnesota Heart Health Program.

Taken together, the results presented in this paper paint a picture of modest and time-limited improvements in exposure to coronary heart disease risk-reducing messages and activities and in coronary heart disease risk factors, usually within chance levels, with no evidence of any cumulative effect on the estimated risk of dying from coronary heart disease. There are several competing explanations, however, and these should be considered.

One competing explanation is that the intervention components themselves did not work. However, there is extensive published research from this study and others demonstrating the efficacy of many of those risk reduction strategies, so this alternative is simply not plausible.[25–47] A second explanation is that the intervention program may have focused on the wrong population segments or used the wrong mix of intervention components. The exposure data suggest that the Minnesota Heart Health Program succeeded in distributing its programs and activities evenly across the strata defined by age, gender, and educational attainment, consistent with the population-based approach of the study, and we cannot know whether the program might have been more effective had it been focused on a particular population segment. In terms of program components, the Minnesota Heart Health Program relied on mass media channels to achieve its awareness goals and on face-to-face direct education activities and communitywide events to achieve its goals for behavior change.[54] Unlike today, there was little support when the Minnesota Heart Health Program was conceived in 1978 for policy-level interventions such as restrictions on smoking and more accurate labeling of food products. Although politically more difficult, these interventions may prove to be more effective than the traditional health education strategies of the 1980s, and it may be that a community program built around a combination of public policy initiatives and the more traditional health education strategies could be more effective. A third alternative explanation is that the Minnesota Heart Health Program's working definition of the community may have been too narrow. The Minnesota program attempted to change the health behaviors and risk behaviors of the residents of three specific communities, with no effort or ability to control or change the broader social milieu in which those communities existed. Just as we have learned that it is difficult to change the behavior of individuals without changing the communities in which they live, we may be learning that it is difficult to change the behavior of whole communities without changing their broader social environment as well. A fourth and final alternative explanation is that health promotion activities such as those developed and implemented in the education communities spread rapidly in the region and effectively "contaminated" the comparison communities. The exposure data presented in Figure 2 provide support for this alternative.

The results from the Minnesota Heart Health Program can be compared with those reported by the Stanford Three- and Five-Community Studies and the North Karelia Study. They also observed modest and often variable improvements in coronary heart disease risk factors after a period of communitywide intervention. In the North Karelia Study,[11] improvements in the intervention community were substantial, but similar trends occurred in the comparison area. In the Stanford Five-Community Project, where the secular trends were not as strong, the risk factor changes were larger and frequently significant.[9] Findings in the Stanford Three- and Five-Community Projects were less frequently significant in analyses

conducted at the city-year level, as we have done in this paper; even so, significant intervention effects remained for some outcomes, including their coronary heart disease risk score.[9,12]

Given the general consistency of the Minnesota Heart Health Program results, it is important to consider the implications of these findings for public health practice and public policy. It is tempting to interpret these results as indicating that the intensive communitywide cardiovascular risk reduction program was ineffective, but that would be an oversimplification. Many of the individual components of the Minnesota Heart Health Program intervention have been shown to be effective in earlier studies. Those components, such as risk factor screening and education, classes for smoking cessation, and dietary change, are still recommended for use by individuals, groups, and communities. What we now recognize is that it is far easier to change the risk profiles of the people who participate in those programs than to engage a large enough fraction of the community to change risk profiles of the entire community. □

Acknowledgments

This research was supported by a grant from the National Heart, Lung and Blood Institute (R01 HL 25523).

The authors wish to thank the staff, students, and volunteers who worked on the Minnesota Heart Health Program. Particular thanks to the citizens and leadership of Mankato, Minn; Fargo, ND; Moorhead, Minn; and Bloomington, Minn. We also wish to recognize F. Gerald Kline, PhD (deceased), and Henry L. Taylor, PhD (deceased), for their many contributions. Finally, we wish to thank the members of the MHHP Scientific Advisory Board for their guidance throughout the study: Thomas Cook, PhD, Northwestern University; Leslie Kish, PhD, University of Michigan; John W. Farquhar, MD, Stanford University; Lawrence W. Green, DrPH, University of British Columbia; Lester Breslow, MD, MPH, University of California at Los Angeles; Kenneth J. Rothman, DrPH, EPI Resources; and Lee Stauffer, PhD, MPH, University of Minnesota.

References

1. Inter-Society Commission for Heart Disease Resources. Optimal resources for prevention of atherosclerotic diseases. *Circulation.* 1984;70:153A–205A.
2. Ross R, Glomset JA. The pathogenesis of atherosclerosis (first of two parts). *N Engl J Med.* 1976;295:369–377.
3. Kane JP, Malloy MJ, Ports TA, Phillips NR, Diehl JC, Havel RJ. Regression of coronary atherosclerosis during treatment of familial hypercholesterolemia with combined drug regimens. *JAMA.* 1990;264: 3007–3012.
4. Blackburn H. Trends and determinants of CHD mortality: changes in risk factors and their effects. *Int J Epidemiol.* 1989;18(suppl 1):S210–S215.
5. World Health Organization. *Community Prevention and Control of Cardiovascular Disease.* Geneva, Switzerland: World Health Organization; 1986. Technical report series 732.
6. Jacobs DR Jr, Luepker RV, Mittelmark MB, et al. Community-wide prevention strategies: evaluation design of the Minnesota Heart Health Program. *J Chronic Dis.* 1986;39:775–788.
7. Mittelmark MB, Luepker RV, Jacobs DR Jr, et al. Community-wide prevention of cardiovascular disease: education strategies of the Minnesota Heart Health Program. *Prev Med.* 1986;15:1–17.
8. Elder JP, McGraw SA, Abrams DB, et al. Organizational and community approaches to community-wide prevention of heart disease: the first two years of the Pawtucket Heart Health Program. *Prev Med.* 1986;15: 107–117.
9. Farquhar JW, Fortmann SP, Flora JA, et al. Effects of community-wide education on cardiovascular disease risk factors: the Stanford Five-City Project. *JAMA.* 1990; 264:359–365.
10. GCP Research Group. *German Cardiovascular Prevention Study: Design, Methods, Results.* Bonn-Bad Godesberg, Germany: Scientific Institute of the German Medical Association; 1988. Program report.
11. Puska P, Salonen J, Nissinen A, et al. Change in risk factors for coronary heart disease during 10 years of community intervention programme: North Karelia Project. *Br Med J.* 1983;287:1840–1844.
12. Williams PT, Fortman SP, Farquhar JW, Varady A, Mellen S. A comparison of statistical methods for evaluating risk factor changes in community-based studies: an example from the Stanford Three-Community Study. *J Chronic Dis.* 1981;34: 565–571.
13. Kish L. *Survey Sampling.* New York, NY: John Wiley & Sons; 1965.
14. Luepker RV, Jacobs DR Jr, Gillum RF, Folsom AF, Prineas RJ, Blackburn H. Population risk of cardiovascular disease: the Minnesota Heart Survey. *J Chronic Dis.* 1985;38:671–682.
15. The Lipids Research Clinics Program. The Coronary Primary Prevention Trial: design and implementation. *J Chronic Dis.* 1979;32: 609–631.
16. The Lipid Research Clinics Program. *Manual of Laboratory Operations. Vol 1. Lipid and Lipoprotein Analysis.* Bethesda, Md: National Institutes of Health; 1974. DHEW publication NIH 75–628.
17. McGovern PG, Jacobs DR Jr, Belcher JD, Luepker RV, Sprafka JM, Blackburn H. The effect of laboratory drift of cholesterol measurement on time trends: the Minnesota Heart Health Program (MHHP). *J Am Coll Cardiol.* 1992;19:344A.
18. Pechacek TF, Murray DM, Luepker RV, Mittelmark MB, Johnson CA, Shultz JM. Measurement of adolescent smoking behavior: rationale and methods. *J Behav Med.* 1984;7:123–140.
19. Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. *J Chronic Dis.* 1967;20:511–524.
20. Leaverton PE, Sorlie PD, Kleinman JC, et al. Representativeness of the Framingham risk model for coronary heart disease mortality. *J Chronic Dis.* 1987;40:775–794.
21. Bandura A. *Social Learning Theory.* Englewood Cliffs, NJ: Prentice-Hall; 1977.
22. Hovland CI, Janis IL, Kelley NH. *Communication and Persuasion.* New Haven, Conn: Yale University Press; 1953.
23. McGuire WJ. Persuasion, resistance and attitude change. In: DeSola I, Schramm W, eds. *Handbook of Communication.* Chicago, Ill: Rand-McNally; 1973:216–252.
24. Rothman J. Three models of community organization practice. In: Cox F, Erlich JL, Rothman J, Tropman JE, eds. *Strategies of Community Organization: A Book of Readings.* Itasca, NJ: Peacock; 1970:20–36.
25. Bracht N, Kingsbury L. Community organization principles in health promotion: a five stage model. In: Bracht N, ed. *Health Promotion at the Community Level.* Newbury Park, Calif: Sage; 1990:66–88.
26. Carlaw R, Mittelmark MB, Bracht NF, Luepker RV. Organization for a community cardiovascular health program: experiences from the Minnesota Heart Health Program. *Health Educ Q.* 1989;16:77–90.
27. Weisbrod RR, Pirie PL, Bracht NF. Impact of a community health promotion program on existing organizations: the Minnesota Heart Health Program. *Soc Sci Med.* 1992;34:639–648.
28. Finnegan JR, Loken B, Howard-Pitney B. Using direct mail to bridge "knowledge gaps" in communication about health. *J Direct Marketing.* 1987;1:29–39.
29. Murray DM, Kurth C, Finnegan JR, Admire JB, Luepker RV. Direct mail as a prompt for followup care among persons at risk for hypertension. *Am J Prev Med.* 1989;4:31–35.
30. Pavlik J, Finnegan JR, Strickland D, Salmon CT, Viswanath K, Wackman DB. Increasing public understanding of heart disease: an analysis of data from the Minnesota Heart Health Program. *Health Commun.* 1993;5:1–20.
31. Salmon C, Loken B, Finnegan JR. Use and effectiveness of direct mail in a cardiovascular health campaign. *Eval Health Professions.* 1985;8:438–452.
32. Viswanath K, Finnegan JR, Hannan P, Luepker RV. Health and knowledge gaps: some lessons from the Minnesota Heart Health Program. *Am Behavioral Scientist.* 1991;34:712–726.
33. Bostick RM, Luepker RV, Kofron PM, Pirie PL. Changes in physician practice for the prevention of cardiovascular disease. *Arch Intern Med.* 1991;151:478–484.
34. Mittelmark MB, Luepker RV, Grimm RH Jr, Kottke TE, Blackburn H. Role of physicians in a community-wide program for prevention of cardiovascular disease: the Minnesota Heart Health Program. *Public Health Rep.* 1988;103:360–365.
35. Murray DM, Luepker RV, Pirie PL, et al. Systematic risk factor screening and education: a community-wide approach to prevention of coronary heart disease. *Prev Med.* 1986;15:661–672.
36. Crow RS, Blackburn H, Jacobs DR, et al. Population strategies to enhance physical activity: the Minnesota Heart Health Program. *Acta Med Scand.* 1986;711(suppl):93–112.

37. Finnegan JR, Murray DM, Kurth C, McCarthy P. Measuring and tracking education program implementation: the Minnesota Heart Health Program Experience. *Health Educ Q.* 1989;16:77–90.
38. Forster JL, Jeffery RW, Schmid TL, Kramer FM. Preventing weight gain in adults: a pound of prevention. *Health Psychol.* 1988;17:129–133.
39. Glanz K, Mullis RM. Environmental interventions to promote healthy eating: a review of models, programs, and evidence. *Health Educ Q.* 1988;15:395–415.
40. Jeffery°RW, Hellerstedt WL, Schmid TL. Correspondence programs for smoking cessation and weight control: a comparison of two strategies in the Minnesota Heart Health Program. *Health Psychol.* 1990;9: 585–598.
41. Lando HA, Hellerstedt WL, Pirie PL, McGovern PG. Brief supportive telephone outreach as a recruitment and intervention strategy for smoking cessation. *Am J Public Health.* 1992;82:41–46.
42. Lando HA, Loken B, Howard-Pitney B, Pechacek T. Community impact of a localized smoking cessation contest. *Am J Public Health.* 1990;80:601–603.
43. Mullis RM, Hunt MK, Foster M, et al. Environmental support of healthful food behavior: the Shop Smart for Your Heart Grocery Program. *J Nutr Educ.* 1987;19: 225–228.
44. Murray DM, Kurth C, Mullis RM, Jeffery RW. Cholesterol reduction through low intensity interventions: results from the Minnesota Heart Health Program. *Prev Med.* 1990;19:181–189.
45. Perry CL, Klepp KI, Sillers C. Community-wide strategies for cardiovascular health: the Minnesota Heart Health Program, Youth Program. *Health Educ Res.* 1989;4: 87–101.
46. Perry CL, Klepp KI, Schultz JM. Primary prevention of cardiovascular disease: community-wide strategies for youth. *J Consult Clin Psychol.* 1988;56:358–364.
47. Perry CL, Kelder SH, Murray DM, Klepp KI. Community-wide smoking prevention: long-term outcomes of the Minnesota Heart Health Program. *Am J Public Health.* 1992;82:1210–1216.
48. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol.* 1981;114:906–914.
49. Zucker DM. An analysis of variance pitfall: the fixed effects analysis in a nested design. *Educ Psychol Meas.* 1990;50:731–738.
50. Murray DM, Hannan PJ, Jacobs DR, et al. Assessing intervention effects in the Minnesota Heart Health Program. *Am J Epidemiol.* 1994;139:91–103.
51. Hannan PJ, Murray DM, Jacobs DR, McGovern PG. Parameters to aid in the design and analysis of community trials: intraclass correlations from the Minnesota Heart Health Program. *Epidemiology.* 1994; 5:88–95.
52. Johnson CL, Rifkind BM, Sempos CT, et al. Declining serum total cholesterol levels among US adults: the National Health and Nutrition Examination Surveys. *JAMA.* 1993;269:3002–3008.
53. Burke GL, Sprafka JM, Folsom AR, Luepker RV, Norsted SW, Blackburn HB. Trends in CHD mortality, morbidity and risk factor levels from 1960–1986. The Minnesota Heart Survey. *Int J Epidemiol.* 1989;18(suppl 1):S73–S81.
54. Flora JA, Lefebvre RC, Murray DM, et al. A community education monitoring system: methods from the Stanford Five-City Project, the Minnesota Heart Health Program and the Pawtucket Heart Health Program. *Health Educ Res.* 1992;8:81–95.

How Effective Is Drug Abuse Resistance Education? A Meta-Analysis of Project DARE Outcome Evaluations

Susan T. Ennett, PhD, Nancy S. Tobler, MS, PhD, Christopher L. Ringwalt, DrPH, and Robert L. Flewelling, PhD

ABSTRACT

Objectives. Project DARE (Drug Abuse Resistance Education) is the most widely used school-based drug use prevention program in the United States, but the findings of rigorous evaluations of its effectiveness have not been considered collectively.

Methods. We used meta-analytic techniques to review eight methodologically rigorous DARE evaluations. Weighted effect size means for several short-term outcomes also were compared with means reported for other drug use prevention programs.

Results. The DARE effect size for drug use behavior ranged from .00 to .11 across the eight studies; the weighted mean for drug use across studies was .06. For all outcomes considered, the DARE effect size means were substantially smaller than those of programs emphasizing social and general competencies and using interactive teaching strategies.

Conclusions. DARE's short-term effectiveness for reducing or preventing drug use behavior is small and is less than for interactive prevention programs. (*Am J Public Health.* 1994;84:1394–1401)

Introduction

School-based drug use prevention programs have been an integral part of the US antidrug campaign for the past two decades.[1,2] Although programs have proliferated, none is more prevalent than Project DARE (Drug Abuse Resistance Education).[3] Created in 1983 by the Los Angeles Police Department and the Los Angeles Unified School District, DARE uses specially trained law enforcement officers to teach a drug use prevention curriculum in elementary schools[4] and, more recently, in junior and senior high schools. Since its inception, DARE has been adopted by approximately 50% of local school districts nationwide, and it continues to spread rapidly.[3] DARE is the only drug use prevention program specifically named in the 1986 Drug-Free Schools and Communities Act. Some 10% of the Drug-Free Schools and Communities Act governors' funds, which are 30% of the funds available each fiscal year for state and local programs, are set aside for programs "such as Project Drug Abuse Resistance Education,"[5] amounting to much of the program's public funding.

Given its widespread use and the considerable investment of government dollars, school time, and law enforcement effort, it is important to know whether DARE is an effective drug use prevention program. That is, to what extent does DARE meet its curriculum objectives, most prominently "to keep kids off drugs"?

DARE's core curriculum, offered to pupils in the last grades of elementary school, is the heart of DARE's program and the focus of this study. We evaluate here the core curriculum's short-term effectiveness by using meta-analytic techniques to integrate the evaluation findings of several studies.[6,7] We searched for all DARE evaluations, both published and unpublished, conducted over the past 10 years and selected for further review those studies that met specified methodological criteria. We calculated effect sizes as a method for establishing a comparable effectiveness measure across studies.[7–9] In addition, to put DARE in the context of other school-based drug use prevention programs, we compared the average magnitude of the DARE effect sizes with those of other programs that target young people of a similar age.

DARE's Core Curriculum

The DARE core curriculum's 17 lessons, usually offered once a week for 45 to 60 minutes, focus on teaching pupils the skills needed to recognize and resist social pressures to use drugs.[4] In addition, lessons focus on providing information about drugs, teaching decision-making skills, building self-esteem, and choosing healthy alternatives to drug use.[4] DARE officers use teaching strategies, such as lectures, group discussions, question-and-

Susan T. Ennett, Christopher L. Ringwalt, and Robert L. Flewelling are with the Research Triangle Institute, Research Triangle Park, NC. Nancy S. Tobler is with the State University of New York, at Albany, NY.

Requests for reprints should be sent to Susan T. Ennett, PhD, Center for Social Research and Policy Analysis, Research Triangle Institute, PO Box 12194, Research Triangle Park, NC 27709-2194.

This paper was accepted February 15, 1994.

Note. The views expressed here are the authors' and do not necessarily represent the official position of the US Department of Justice or the US Department of Health and Human Services.

answer sessions, audiovisual material, workbook exercises, and role-playing.[4]

The training that DARE officers receive is substantial. They are required to undergo 80 training hours in classroom management, teaching strategies, communication skills, adolescent development, drug information, and curriculum instruction.[4] In addition, DARE officers with classroom experience can undergo further training to qualify as instructors/mentors.[4] These officers monitor the program delivery's integrity and consistency through periodic classroom visits.

Methods

Identification of Evaluations

We attempted to locate all quantitative evaluations of DARE's core curriculum through a survey of DARE's five Regional Training Centers, computerized searches of the published and unpublished literature, and telephone interviews with individuals known to be involved with DARE. Eighteen evaluations in 12 states and one province in Canada were identified. Several evaluations were reported in multiple reports or papers. (See Appendix A for a bibliography of the studies considered.)

Evaluation Selection Criteria

To be selected for this meta-analysis, an evaluation must have met the following criteria: (1) use of a control or comparison group; (2) pretest-posttest design or posttest only with random assignment; and (3) use of reliably operationalized quantitative outcome measures. Quasi-experimental studies were excluded if they did not control for preexisting differences on measured outcomes with either change scores or covariance-adjusted means.[10] In addition, to ensure comparability, we focused on results based only on immediate posttest. Because only four evaluation studies were long term (two of which were compromised by severe control group attrition or contamination), we were unable to adequately assess longer-term DARE effects.

We examined several other methodological features, such as the correspondence between the unit of assignment and analysis, the use of a panel design, matching of schools in the intervention and control conditions, and attrition rates. Although these factors were considered in assessing the studies' overall methodological rigor, we did not eliminate evaluations on the basis of these criteria.

Data Analysis

For each study, we calculated an effect size to quantify the magnitude of DARE's effectiveness with respect to each of six outcomes that reflect the DARE curriculum's aims. An effect size is defined as the difference between the intervention and the control group means for each outcome measure, standardized by dividing by the pooled standard deviation [effect size = $\text{mean}_I - \text{mean}_C$/SD].[7-9] If means and standard deviations were not available, we calculated effect sizes using formulas developed to convert other test statistics and percentages to effect sizes.[9] In all cases, we used statistics reflecting covariance-adjusted means, with pretest values as covariates rather than unadjusted means so that any differences between the comparison groups before the intervention would not be reflected in the effect sizes.[10]

The six outcome measure classes include knowledge about drugs, attitudes about drug use, social skills, self-esteem, attitude toward police, and drug use. Some studies did not include all six, and some outcomes were measured by more than one indicator. When multiple indicators were used (e.g., two measures of social skills), we calculated separate effect sizes and then averaged them.[6,10] This procedure yielded one effect size per study for each measured outcome type. In the one study that reported only that a measured outcome was not statistically significant (and did not provide any further statistics), we assigned a zero value to that effect size.[10] To calculate effect sizes for drug use, we considered only alcohol, tobacco, and marijuana use; we averaged effect sizes across these substances. In a supplementary analysis, we considered use of these substances separately. The prevalence of other drugs, such as cocaine, was too small to produce meaningful effects.

In addition to calculating one effect size per outcome per study, we calculated the weighted mean effect size and 95% confidence interval (CI) for each outcome type across programs. The weighted mean is computed by weighting each effect size by the inverse of its variance, which is a reflection of the sample size.[8,9] The effect size estimates from larger studies are generally more precise than those from smaller studies.[8] Hence, the weighted mean provides a less biased estimate than the simple, unweighted mean because estimates from larger samples are given more weight. The 95% CI indicates the estimated effect size's accuracy or reliability and is calculated by adding to or subtracting from the mean 1.96 multiplied by the square root of 1 divided by the sum of the study weights.[8]

TABLE 1—DARE Evaluation Studies Selected for Review

Location	References[a]
British Columbia (BC)	Walker 1990
Hawaii (HI)	Manos, Kameoka, and Tanji 1986
Illinois (IL)	Ennett et al. 1994 (in press)
Kentucky-A (KY-A)	Clayton et al. 1991a,1991b
Kentucky-B (KY-B)	Faine and Bohlander 1988, 1989
Minnesota (MN)	McCormick and McCormick 1992
North Carolina (NC)	Ringwalt, Ennett, and Holt 1991
South Carolina (SC)	Harmon 1993

[a]See Appendix A for full references.

Comparison of DARE with Other Drug Use Prevention Programs

For comparison with DARE, we used the effect sizes reported in Tobler's meta-analysis of school-based drug use prevention programs.[10] To allow the most appropriate comparisons with DARE effect sizes, we obtained Tobler's results for only those programs (excluding DARE) aimed at upper elementary school pupils. These programs are a subset of 25 from the 114 programs in Tobler's meta-analysis, whose studies are referenced in Appendix B.

We selected this meta-analysis for comparison because of its greater similarity to ours than other meta-analyses of drug use prevention programs.[11-14] Tobler's studies met the same methodological standards that we used for the DARE studies. The only differences were that Tobler excluded studies that did not measure drug use and considered results from later posttests, whereas we considered only immediate posttest results. Neither of these differences, however, should seriously compromise the comparison.

The evaluation studies included in Tobler's meta-analysis are classified into

TABLE 2—Sample and Methodological Characteristics of the DARE Evaluations (n = 8)

Study	Schools, n	Subjects, n	Research Design	Matching	Unit of Analysis	Pretest Equivalency[a]	Scale Reliabilities	Attrition
BC	11	D = 287 C = 175	Quasi, cross-sectional	Yes	Individual	Yes	No	Not applicable
HI	26	D = 1574 C = 435	Quasi, panel	No	Individual	No	No	No
IL	36	D = 715 C = 608	Experimental/quasi, panel	Yes	School based	Yes	Yes	Yes[b]
KY-A	31	D = 1438 C = 487	Experimental, panel	No	Individual	Yes	Yes	Yes[b]
KY-B	16	D = 451 C = 332	Quasi, panel	Yes	Individual	Yes	Yes	No
MN	63	D = 453 C = 490	Quasi, panel	No	Individual	Yes	Yes	Yes[c]
NC	20	D = 685 C = 585	Experimental, panel	No	School based	Yes	Yes	Yes[b]
SC	11	D = 295 C = 307	Quasi, panel	Yes	Individual	Yes	Yes	Yes[c]

Note. See Table 1 for information on study locations and references. D = DARE; C = comparison.
[a]Pretest equivalency on demographic variables assessed and controlled if necessary.
[b]Attrition rates reported and differential attrition across experimental conditions analyzed.
[c]Attrition rates reported only.

TABLE 3—Unweighted Effect Sizes Associated with Eight DARE Evaluations

Study	Knowledge	Attitudes about Drugs	Social Skills	Self-Esteem	Attitude toward Police	Drug Use[a]
BC	.68	.00	...	...	...	.02
HI	...	.07	.34	...	...	...
IL	...	.03	.15	.15	.12	.05
KY-A	...	.11	.10	.07	...	.00
KY-B	.58	.19	.30	.14	.27	...
MN	.19	.06	.08	−.03	.05	...
NC	...	.19	.17	.00	...	.11
SC	...	.32	.19	.06	.08	.10

Note. See Table 1 for information on study locations and references.
[a]Limited to alcohol, tobacco, and marijuana.

two broad categories based on the programs' content and process. Process describes the teaching approach (how the content is delivered). Programs classified by Tobler as "noninteractive" emphasize intrapersonal factors, such as knowledge gain and affective growth, and are primarily delivered by an expert. "Interactive" programs emphasize interpersonal factors by focusing on social skills and general social competencies and by using interactive teaching strategies, particularly peer to peer. Consistent with other meta-analyses showing that programs emphasizing social skills tend to be the most successful,[11–13,15] interactive programs produced larger effect sizes than noninteractive programs. We compared DARE with both categories of programs.

Results

Characteristics of Evaluations

Of the original 18 studies, 8 met the criteria for inclusion. One additional study met the methodological criteria but did not administer the first posttest until 1 year after DARE implementation; therefore, it could not be included in our analysis of immediate effects.[16,17] The location and primary reference for each evaluation are shown in Table 1, and study characteristics are summarized in Table 2.

Each evaluation represents a state or local effort. The number of student subjects in all studies was large, each study comprising at least 10 schools with approximately 500 to 2000 students. Although demographic information was not given for three studies, the remaining five studies in the sample primarily consisted of White subjects.

Assignment of DARE to intervention and control groups was by school for all eight studies. In one study, DARE also was assigned by classroom in certain schools.[18] Because of potential contamination in this study of the control group classrooms by their close proximity to DARE classes, we eliminated these control classrooms; only control schools with no DARE classes were included. Two studies used a true experimental design in which schools were randomly assigned to DARE and control conditions; a third study used random assignment for two thirds of the schools. The remaining five evaluations used a nonequivalent control group quasi-experimental design.

Because there were relatively few sampling units across studies—ranging from 11 to 63 schools, with all except one study involving fewer than 40 schools—it is unlikely that equivalence between

groups was obtained without prior matching or blocking of schools, even with randomization. Only half the studies matched comparison schools on selected demographic characteristics. Most studies (75%), however, assessed the equivalency of the comparison groups at pretest and made adjustments for pretest differences on demographic characteristics. All studies adjusted for pretest differences on outcome measures.

All but one study used a panel design that matched subjects from pretest to posttest with a unique identification code.

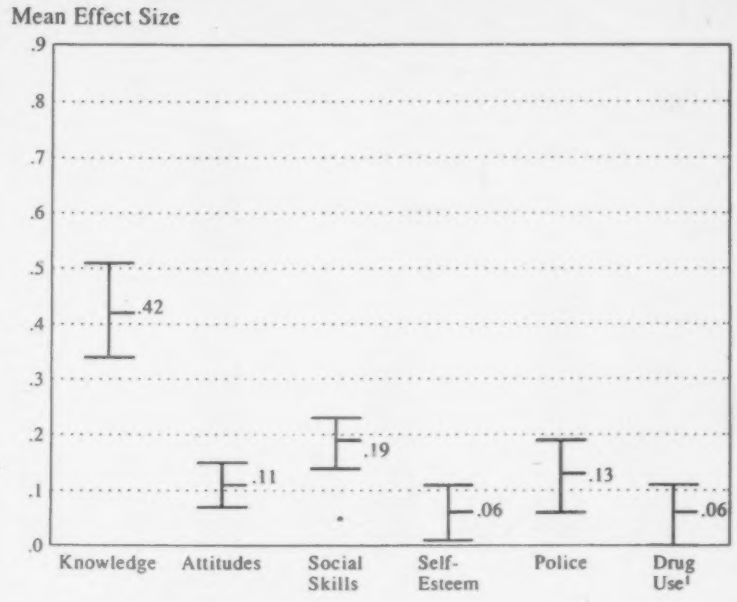
Outcome measures used in the DARE evaluations were based on responses to self-administered questionnaires. Seven studies used standardized scales or revised existing measures; six studies reported generally high scale reliabilities (usually Cronbach's alpha). Validity information, however, was rarely reported, and no study used either a biochemical indicator or "bogus pipeline" technique to validate drug use self-reports.[19]

Most studies (75%) did not use a data analysis strategy appropriate to the unit of assignment. Because schools, not students, were assigned to DARE and control conditions, it would have been appropriate to analyze the data by schools with subjects' data aggregated within each school or to use a hierarchical analysis strategy in which subjects are nested within schools.[20,21] Six studies ignored schools altogether and analyzed individual subjects' data, thereby violating the statistical assumption of independence of observations. Ignoring schools as a unit of analysis results in a positive bias toward finding statistically significant program effects.[21] This bias may be reflected in CIs reported for each outcome's weighted mean effect size.

Five studies reported generally small attrition rates. None of the three studies that analyzed attrition found that rates differed significantly across experimental and control conditions. In addition, subjects absent from the posttest were not more likely to be drug users or at risk for drug use. Although attrition usually is greater among drug users,[22] given the sample's young age (when school dropout is unlikely and drug use prevalence is low), these results are not surprising.

DARE Effect Sizes

Study effect sizes are shown in Table 3. In general, the largest effect sizes are for knowledge and social skills; the smallest are for drug use.



[1]Drug use includes alcohol, tobacco, and marijuana.

FIGURE 1—Magnitude of DARE's weighted mean effect size (and 95% CI), by outcome measure.

Figure 1 shows the mean weighted effect size and 95% CI for each outcome based on the eight studies combined. The largest mean effect size is for knowledge (.42), followed by social skills (.19), attitude toward the police (.13), attitudes about drug use (.11), self-esteem (.06), and drug behavior (.06). The effect sizes for knowledge, social skills, attitude toward the police, attitudes about drug use, and self-esteem are statistically significant. The CI for the mean drug use effect size overlaps with zero (i.e., it is not significantly different from zero).

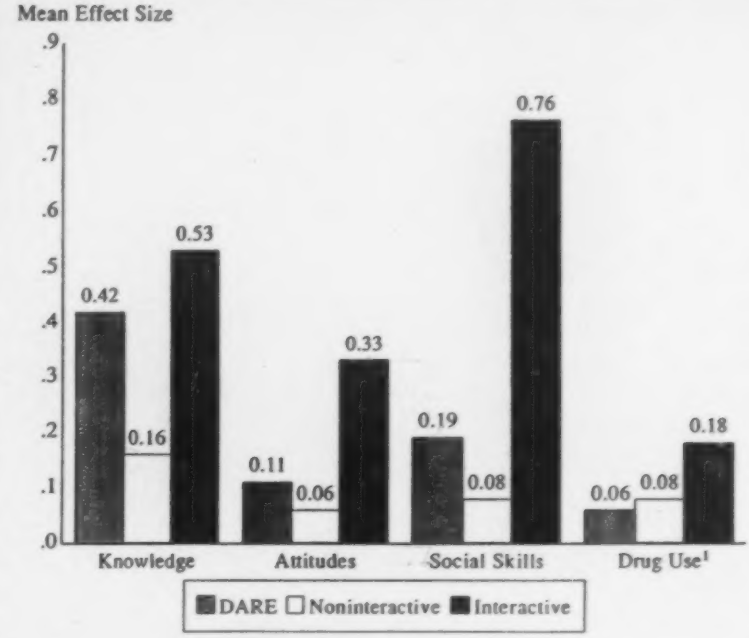
Because averaging alcohol, tobacco, and marijuana use for the drug use effect size could obscure substantial differences among the substances, we calculated DARE's mean weighted effect sizes separately for these substances. The weighted mean effect size for alcohol use is .06 (95% CI = .00, .12); for tobacco use, .08 (95% CI = .02, .14); and for marijuana use, −.01 (95% CI = −.09, .07). Only the mean for tobacco use is statistically significant.

Mean Effect Sizes for DARE vs Other Drug Use Prevention Programs

We compared by type of outcome the mean weighted DARE effect size with the mean weighted effect size for noninteractive (n = 9) and interactive (n = 16) programs; effect sizes for the comparison programs are derived from Tobler.[10] The comparison programs target youth of the same grade range targeted by DARE. The outcomes assessed by both DARE and the comparison programs are knowledge, attitudes, social skills, and drug use behavior.

Across the four outcome domains, DARE's effect sizes are smaller than those for interactive programs (Figure 2). Most notable are DARE's effect sizes for drug use and social skills; neither effect size (.06 and .19, respectively) is more than a third of the comparable effect sizes for interactive programs (.18 and .75, respectively). DARE's effect size for drug use is only slightly smaller than the noninteractive programs' effect size. DARE's effect sizes for knowledge, attitudes, and social skills, however, are larger than those for noninteractive programs.

Comparison of effect sizes separately for alcohol, tobacco, and marijuana use shows that DARE's effect sizes are smaller than those for interactive programs (Figure 3). Except for tobacco use,



Note. Comparison programs selected from Tobler.[10]
[1]Drug use includes alcohol, tobacco, and marijuana.

FIGURE 2— Weighted mean effect size, by outcome, for DARE and other drug use prevention programs.

they also are smaller than those for noninteractive programs.

Discussion

The results of this meta-analysis suggest that DARE's core curriculum effect on drug use relative to whatever drug education (if any) was offered in the control schools is slight and, except for tobacco use, is not statistically significant. Across the studies, none of the average drug use effect sizes exceeded .11. Review of several meta-analyses of adolescent drug use prevention programs suggests that effect sizes of this magnitude are small.[10–14]

The small magnitude of DARE's effectiveness on drug use behavior may partially reflect the relatively low frequency of drug use by the elementary school pupils targeted by DARE's core curriculum. However, comparison of the DARE effect sizes with those of other school-based drug use prevention programs for same-age adolescents suggests that greater effectiveness is possible with early adolescents. Compared with the programs classified by Tobler as interactive, DARE's effect sizes for alcohol, tobacco, and marijuana use, both collectively and individually, are substantially less.[10] Except for tobacco use, they also are less than the drug use effect sizes for more traditional, noninteractive programs.

It has been suggested that DARE may have delayed effects on drug use behavior once pupils reach higher grades.[23,24] Longer-term follow-up studies are needed to test this possibility. Only four reviewed studies administered multiple posttests, and for two of these the results from some later posttests are uninterpretable. However, based on two experimental studies for which reliable information 1 and 2 years after implementation is available, there is no evidence that DARE's effects are activated when subjects are older.[25,26] Most long-term evaluations of drug use prevention programs have shown that curriculum effects decay rather than appear or increase with time.[27,28]

DARE's immediate effects on outcomes other than drug use were somewhat larger (especially for knowledge) and were statistically significant. These effect sizes, however, also were less than the comparable effect sizes for same-age interactive programs. That DARE's effect sizes for knowledge, attitudes, and skills were greater in magnitude than those of noninteractive programs may not be particularly meaningful because many of these types of programs, such as programs using "scare tactics" or emphasizing factual knowledge about drug use, have been discredited as unsuccessful.[29,30]

Comparison of DARE's core curriculum content with the interactive and noninteractive programs' curricula may partially explain the relative differences in effect sizes among these programs. Interactive programs tend to emphasize developing drug-specific social skills and more general social competencies, whereas noninteractive programs focus largely on intrapersonal factors. Because DARE has features of both interactive and noninteractive programs, it is perhaps not surprising that the effect sizes we reported should fall somewhere in between. Perhaps greater emphasis in the DARE core curriculum on social competencies and less emphasis on affective factors might result in effect sizes nearer to those reported for interactive programs. However, it is difficult to speculate on the effect of adding or subtracting particular lessons to or from DARE's curriculum. Most school-based prevention program evaluations have assessed the effectiveness of an overall program rather than various program components or combinations of components.

Who teaches DARE and how it is taught may provide other possible explanations for DARE's limited effectiveness. Despite the extensive DARE training received by law enforcement officers, they may not be as well equipped to lead the curriculum as teachers. No studies have been reported in which the DARE curriculum was offered by anyone other than a police officer; results from such a study might suggest whether teachers produce better (or worse) outcomes among pupils.

Regardless of curriculum leader, however, the generally more traditional teaching style used by DARE has not been shown to be as effective as an interactive teaching mode.[10,14] Although some activities encourage pupil interaction, the curriculum relies heavily on the officer as expert and makes frequent use of lectures and question-and-answer sessions between the officer and pupils. In fact, it is in teaching style, not curriculum content, that DARE most differs from the interactive programs examined by Tobler.

The DARE core curriculum recently was modified to introduce more participatory activities, which may lead to greater program effectiveness.

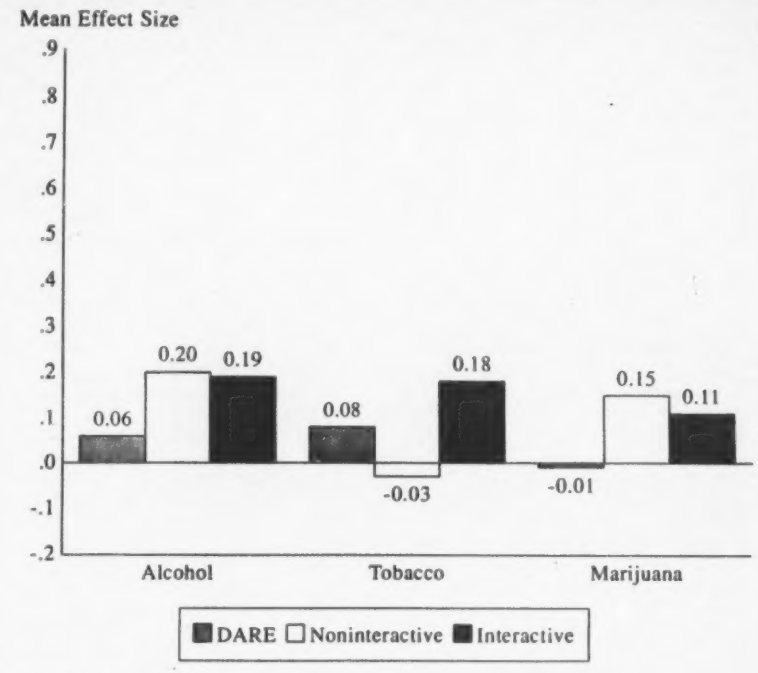
Several limitations should be considered in evaluating our findings. The number of evaluations reviewed (eight) is not large when compared with the vast number of sites where DARE has been implemented. The consistency of results across studies, however, suggests that the results are likely to be representative of DARE's core curriculum. Even so, we would have preferred a full set of eight effect sizes for each outcome.

It is possible that the effect sizes for the DARE studies may have been attenuated compared with the drug use prevention programs reviewed by Tobler because the control groups were not pure "no treatment" groups. As documented by Tobler, effect sizes are lower when the control group receives some sort of drug education.[10,14] The DARE evaluations generally lacked information on alternative treatments received by the control groups, but it is likely that most control groups received some drug education because the studies occurred after the 1986 Drug-Free Schools and Communities Act. However, approximately half (54%) of the programs reviewed by Tobler also were conducted between 1986 and 1990, suggesting that they may suffer from the same effect.[10]

Most of the drug use prevention programs evaluated by Tobler were university research-based evaluation studies, whereas DARE is a commercially available curriculum. Although the magnitude of the resources invested in DARE is considerable, the intensity of effort devoted to smaller-scale programs may be greater. Some diminished effectiveness is perhaps inevitable once programs are widely marketed.

Although we found limited immediate core curriculum effects, some features of DARE may be more effective, such as the middle school curriculum. In addition, DARE's cumulative effects may be greater in school districts where all DARE curricula for younger and older students are in place. Other DARE outcomes, such as its impact on community law enforcement relations, also may yield important benefits. However, due to the absence of evaluation studies, consideration of these features is beyond this study's scope.

DARE's limited influence on adolescent drug use behavior contrasts with the program's popularity and prevalence. An important implication is that DARE could be taking the place of other, more beneficial drug use curricula that adolescents could be receiving. At the same time, expectations concerning the effectiveness of any school-based curriculum, including DARE, in changing adolescent drug use behavior should not be overstated.[31] □



Note. Comparison programs selected from Tobler.[10]

FIGURE 3—Weighted mean effect size, by drug, for DARE and other drug use prevention programs.

Acknowledgments

This research was supported by awards from the National Institute of Justice, Office of Justice Programs, US Department of Justice (91-DD-CX-K053) and the National Institute on Drug Abuse, US Department of Health and Human Services (5 R01 DA07037).

We thank Susan L. Bailey, Karl E. Bauman, George H. Dunteman, Robert L. Hubbard, Ronald W. Johnson, and J. Valley Rachal for their thoughtful comments on an earlier draft of this paper. In addition, we acknowledge the support and assistance of Jody M. Greene on this project and the editorial assistance of Beryl C. Pittman, Linda B. Barker, and Richard S. Straw.

References

1. Glynn TJ. Essential elements of school-based smoking prevention program. *J Sch Health.* 1989;59:181–188.
2. *Healthy People 2000.* Washington, DC: US Dept of Health and Human Services; 1988. DHHS publication no. PHS 91-50212.
3. Ringwalt CL, Greene JM. Results of school districts' drug prevention coordinator's survey. Presented at the Alcohol, Tobacco, and Other Drugs Conference on Evaluating School-Linked Prevention Strategies; March 1993; San Diego, Calif.
4. *Implementing Project DARE: Drug Abuse Resistance Education.* Washington, DC: Bureau of Justice Assistance; 1988.
5. Drug-Free Schools and Community Act. Pub L No. 101-647, §5122(e)(1).
6. Bangert-Drowns RL. Review of developments in meta-analytic method. *Psychol Bull.* 1986;99:388–399.
7. Rosenthal R. *Meta-Analytic Procedures for Social Research.* Applied Social Research Methods Series. Vol 6. Newbury Park, Calif: Sage Publications; 1991.
8. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis.* New York, NY: Academic Press; 1985.
9. Perry PD, Tobler NS. Meta-analysis of adolescent drug prevention programs: final report. In: *Manual for Effect Size Calculation: Formula Used in the Meta-Analysis of Adolescent Drug Prevention Programs.* Rockville, Md: National Institute on Drug Abuse; 1992: appendix 3.
10. Tobler NS. *Meta-Analysis of Adolescent Drug Prevention Programs: Final Report.*

Rockville, Md: National Institute on Drug Abuse; 1992.
11. Bangert-Drowns RL. The effects of school-based substance abuse education: a meta-analysis. *J Drug Educ.* 1988;18:243–264.
12. Bruvold WH. A meta-analysis of adolescent smoking prevention programs. *Am J Public Health.* 1993;83:872–880.
13. Bruvold WH, Rundall TG. A meta-analysis and theoretical review of school-based tobacco and alcohol intervention programs. *Psychol Health.* 1988;2:53–78.
14. Tobler NS. Meta-analysis of 143 adolescent drug prevention programs: quantitative outcome results of program participants compared to a control or comparison group. *J Drug Issues.* 1986;16:537–567.
15. Hansen WB. School-based substance abuse prevention: a review of the state of the art in curriculum, 1980–1990. *Health Educ Res.* 1992;7:403–430.
16. Nyre GF, Rose C. *Drug Abuse Resistance Education (DARE) Longitudinal Evaluation Annual Report, January 1987.* Los Angeles, Calif: Evaluation and Training Institute; 1987. Unpublished report.
17. Nyre GF, Rose C, Bolus RE. *Drug Abuse Resistance Education (DARE) Longitudinal Evaluation Annual Report, August 1987.* Los Angeles, Calif: Evaluation and Training Institute; 1987. Unpublished report.
18. Manos MJ, Kameoka KY, Tanji JH. *Evaluation of Honolulu Police Department's Drug Abuse Resistance Education Program.* Honolulu, Hawaii: University of Hawaii-Manoa, School of Social Work, Youth Development and Research Center; 1986. Unpublished report.
19. Bauman KE, Dent CW. Influence of an objective measure on the self-reports of behavior. *J Appl Psychol.* 1982;67:623–628.
20. Moskowitz JM. Why reports of outcome evaluations are often biased or uninterpretable. *Eval Progr Plan.* 1993;16:1–9.
21. Murray DM, Hannan PJ. Planning for the appropriate analysis in school-based drug use prevention studies. *J Consult Clin Psychol.* 1990;58:458–468.
22. Biglan A, Ary DV. Methodological issues in research on smoking prevention. In: Bell CS, Battjes R, eds. *Prevention Research: Deterring Drug Abuse Among Children and Adolescents.* Rockville, Md: National Institute on Drug Abuse; 1985:170–195.
23. Clayton RR, Cattarello A, Day LE, Walden KP. Persuasive communication and drug abuse prevention: an evaluation of the DARE program. In: Donohew L, Sypher H, Bukoski W, eds. *Persuasive Communication and Drug Abuse Prevention.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1991:295–313.
24. Ringwalt CL, Ennett ST, Holt KD. An outcome evaluation of Project DARE (Drug Abuse Resistance Education). *Health Educ Res.* 1991;6:327–337.
25. Clayton RR, Cattarello A, Walden KP. Sensation seeking as a potential mediating variable for school-based prevention intervention: a two-year follow-up of DARE. *Health Commun.* 1991;3:229–239.
26. Ennett ST, Rosenbaum DP, Flewelling RL, Bieler GS, Ringwalt CL, Bailey SL. Long-term evaluation of Drug Abuse Resistance Education. *Addict Behav.* 1994;19:113–125.
27. Ellickson PL, Bell RM, McGuigan K. Preventing adolescent drug use: long-term results of a junior high program. *Am J Public Health.* 1993;83:856–861.
28. Murray DM, Pirie P, Luepker RV, Pallonen U. Five- and six-year follow-up results from four seventh-grade smoking prevention strategies. *J Behav Med.* 1989;12:207–218.
29. Botvin GJ. Substance abuse prevention: theory, practice, and effectiveness. In: Tonry M, Wilson JQ, eds. *Drugs and Crime.* Chicago, Ill: University of Chicago Press; 1990:461–519.
30. Moskowitz J. Preventing adolescent substance abuse through education. In: Glynn T, Leukefeld CG, Ludford JP, eds. *Preventing Adolescent Drug Abuse: Intervention Strategies.* Rockville, Md: National Institute on Drug Abuse; 1983:233–249.
31. Dryfoo JG. Preventing substance abuse: rethinking strategies. *Am J Public Health.* 1993;83:793–795.

APPENDIX A—Bibliography of DARE Evaluations

Aniskiewicz RE, Wysong EE. *Project DARE Evaluation Report: Kokomo Schools Spring, 1987.* Kokomo, Ind: Indiana University–Kokomo; 1987. Unpublished report.

Aniskiewicz R, Wysong E. Evaluating DARE: drug education and the multiple meanings of success. *Policy Stud Rev.* 1990;9:727–747.

Anonymous. *Project DARE Evaluation, Spring, 1987.* Pittsburgh, Pa; 1987. Unpublished paper.

Becker HR, Agopian MW, Yeh S. Impact evaluation of Drug Abuse Resistance Education (DARE). *J Drug Educ.* 1990;22:283–291.

Clayton RR, Cattarello A, Day LE, Walden KP. Persuasive communication and drug abuse prevention: an evaluation of the DARE program. In: Donohew L, Sypher H, Bukoski W, eds. *Persuasive Communication and Drug Abuse Prevention.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1991a:295–313.

Clayton RR, Cattarello A, Walden KP. Sensation seeking as a potential mediating variable for school-based prevention intervention: a two-year follow-up of DARE. *Health Commun.* 1991b;3:229–239.

DeJong W. A short-term evaluation of Project DARE (Drug Abuse Resistance Education): preliminary indicators of effectiveness. *J Drug Educ.* 1987;17:279–294.

Dukes R, Matthews S. *Evaluation of the Colorado Springs Citywide DARE Program.* Colorado Springs, Colo: University of Colorado, Center for Social Science Research; 1991. Unpublished report.

Earle RB, Garner J, Phillips N. *Evaluation of the Illinois State Police Pilot DARE Program.* Springfield, Ill: A.H. Training and Development Systems, Inc; 1987. Unpublished report.

Ennett ST, Rosenbaum DP, Flewelling RL, Bieler GS, Ringwalt CR, Bailey SL. Long-term evaluation of Drug Abuse Resistance Education. *Addict Behav.* 1994;19:113–125.

Evaluation and Training Institute. *DARE Evaluation Report for 1985–1989.* Los Angeles, Calif: Evaluation and Training Institute; 1990. Unpublished report.

Faine JR, Bohlander E. *Drug Abuse Resistance Education: An Assessment of the 1987–88 Kentucky State Police DARE Program.* Bowling Green, Ky: Western Kentucky University, Social Research Laboratory; 1988. Unpublished report.

Faine JR, Bohlander E. *DARE in Kentucky Schools 1988–89.* Bowling Green, Ky: Western Kentucky University, Social Research Laboratory; 1989. Unpublished report.

Harmon MA. Reducing the risk of drug involvement among early adolescents: an evaluation of Drug Abuse Resistance Education (DARE). *Eval Rev.* 1993;17:221–239.

Kethineni S, Leamy DA, Guyon L. *Evaluation of the Drug Abuse Resistance Education Program in Illinois: Preliminary Report.* Normal, Ill: Illinois State University, Department of Criminal Justice Sciences; 1991. Unpublished report.

McCormick, FC, McCormick ER. *An Evaluation of the Third Year Drug Abuse Resistance Education (DARE) Program in Saint Paul.* Saint Paul, Minn: Educational Operations Concepts, Inc; 1992. Unpublished report.

McDonald RM, Towberman DB, Hague JL. *Volume II: 1989 Impact Assessment of Drug Abuse Resistance Education in the Commonwealth of Virginia.* Richmond, Va: Virginia Commonwealth University, Institute for Research in Justice and Risk Administration; 1990. Unpublished report.

Manos MJ, Kameoka KY, Tanji JH. *Evaluation of Honolulu Police Department's Drug Abuse Resistance Education Program.* Honolulu, Hawaii: University of Hawaii–Manoa, School of Social Work, Youth Development and Research Center; 1986. Unpublished report.

APPENDIX A—Continued

Nyre GF, Rose C. *Drug Abuse Resistance Education (DARE) Longitudinal Evaluation Annual Report, January 1987.* Los Angeles, Calif: Evaluation and Training Institute; 1987. Unpublished report.

Nyre GF, Rose C, Bolus RE. *Drug Abuse Resistance Education (DARE) Longitudinal Evaluation Annual Report, August 1987.* Los Angeles, Calif: Evaluation and Training Institute; 1987. Unpublished report.

Ringwalt C, Ennett ST, Holt KD. An outcome evaluation of Project DARE (Drug Abuse Resistance Education). *Health Educ Res.* 1991;6:327–337.

Walker S. *The Victoria Police Department Drug Abuse Resistance Education Programme (D.A.R.E.) Programme Evaluation Report #2 (1990).* Victoria, British Columbia: The Ministry of the Solicitor-General, Federal Government of Canada; 1990. Unpublished report.

APPENDIX B—Bibliography of Comparison Program Evaluations

Allison K, Silver G. Dignam C. Effects on students of teacher training in use of a drug education curriculum. *J Drug Educ.* 1990;20:31–46.

Dielman T, Shope J, Butchart A, Campanelli P. Prevention of adolescent alcohol misuse: an elementary school program. *J Pediatr Psychol.* 1986;11:259–281.

Dielman T, Shope J, Campanelli P, Butchart A. Elementary school-based prevention of adolescent alcohol misuse. *Pediatrician: Int J Child Adolesc Health.* 1987;14:70–76.

Dielman T, Shope J, Leech S, Butchart A. Differential effectiveness of an elementary school-based alcohol misuse prevention program. *J Sch Health.* 1989;59:255–262.

Dubois R, Hostelter M, Tosti-Vasey J, Swisher J. *Program Report and Evaluation: 1988–1989 School Year.* Philadelphia, Pa: Corporate Alliance for Drug Education; 1989. Unpublished report.

Flay B, Koepke D, Thomson S, Santi S, Best J, Brown S. Six-year follow-up of the first Waterloo school smoking prevention trial. *Am J Public Health.* 1989;79:1371–1376.

Flay B, Ryan K, Best J, et al. Cigarette smoking: why young people do it and ways of preventing it. In: McGrath P, Firestone P, eds. *Pediatric and Adolescent Behavioral Medicine.* New York, NY: Springer-Verlag, 1983:132–183.

Flay B, Ryan K, Best J, et al. Are social psychological smoking prevention programs effective? The Waterloo study. *J Behav Med.* 1985;8:37–59.

Gersick K, Grady K, Snow D. Social-cognitive development with sixth graders and its initial impact on substance use. *J Drug Educ.* 1988;18:55–70.

Gilchrist L, Schinke S, Trimble J, Cvetkovich G. Skills enhancement to prevent substance abuse among American Indian adolescents. *Int J Addict.* 1987;22:869–879.

Johnson C, Graham J, Hansen W, Flay B, McGuigan K, Gee M. *Project Smart After Three Years: An Assessment of Sixth-Grade and Multiple-Grade Implementations.* Pasadena, Calif: University of Southern California, Institute for Prevention Research; 1987. Unpublished report.

McAlister A. Approaches to primary prevention. Presented at the National Academy of Science National Research Council Conference on Alcohol and Public Policy; May 1983; Washington, DC.

Moskowitz J, Malvin J, Schaeffer G, Schaps E. Evaluation of an affective development teacher training program. *J Primary Prev.* 1984;4:150–162.

Sarvela P. Early adolescent substance use in rural northern Michigan and northeastern Wisconsin. In: *Dissertation Abstracts International.* 1984:46 01, 2000A. University Microfilms No. 84-22, 327.

Sarvela P, McClendon E. An impact evaluation of a rural youth drug education program. *J Drug Educ.* 1987; 17:213–231.

Schaeffer G, Moskowitz J, Malvin J, Schaps E. *The Effects of a Classroom Management Teacher Training Program on First-Grade Students: One Year Follow-Up.* Napa, Calif: The Napa Project, Pacific Institute for Research; 1981. Unpublished report.

Schaps E, Moskowitz J, Condon J, Malvin J. A process and outcome evaluation of an affective teacher training primary prevention program. *J Alcohol Drug Educ.* 1984;29:35–64.

Schinke S, Bebel M, Orlandi M, Botvin G. Prevention strategies for vulnerable pupils: school social work practices to prevent substance abuse. *Urban Educ.* 1988;22:510–519.

Schinke SP, Blythe BJ. Cognitive-behavioral prevention of children's smoking. *Child Behav Ther.* 1981;3:25–42.

Schinke S, Gilchrist L. Primary prevention of tobacco smoking. *J Sch Health.* 1983;53:416–419.

Schinke S, Gilchrist L, Schilling R, Snow W, Bobo J. Skills methods to prevent smoking. *Health Educ Q.* 1986;13:23–27.

Schinke S, Gilchrist L, Snow W. Skills interventions to prevent cigarette smoking among adolescents. *Am J Public Health.* 1985;75:665–667.

Schinke S, Gilchrist L, Snow W, Schilling R. Skills-building methods to prevent smoking by adolescents. *J. Adolesc Health Care.* 1985;6:439–444.

Shope J, Dielman T, Leech S. An elementary school-based alcohol misuse prevention program: two year follow-up evaluation. Presented at the American Public Health Association Annual Meeting; November 1988; Boston, Mass.

Drinking Behavior in Relation to Cause of Death among US Adults

ABSTRACT

Objectives. Alcohol involvement is common in many fatal injuries. This study examines drinking behavior in a nationally representative sample of US adult decedents aged 25 through 64 years and its association with cause of death.

Methods. Proxy-reported information from the 1986 National Mortality Followback Survey was used to profile the decedents' usual frequency and quantity of drinking. The association of drinking behavior with underlying cause of death was assessed while adjusting for demographic characteristics.

Results. Of the decedents, 17% were daily drinkers, 22% usually consumed five drinks or more per occasion, and 27% were classified as heavier drinkers. Persons who died of injury drank more frequently and heavily than those who died of disease. The adjusted odds ratio of injury's being the underlying cause of death was 1.4 (95% confidence interval [CI] = 1.1, 1.8) for daily drinkers, 1.5 (95% CI = 1.1, 2.0) for those drinking five or more drinks per occasion, and 1.4 (95% CI = 1.1, 1.7) for heavier drinkers.

Conclusions. Daily drinking, binge drinking, and heavier drinking were each associated with an increased likelihood of injury as the underlying cause of death. Persons who were young, male, Native American, or divorced or separated were more likely to drink frequently and heavily. (*Am J Public Health.* 1994;84: 1402–1406)

Guohua Li, MD, DrPH, Gordon S. Smith, MD, MPH, and Susan P. Baker, MPH

Introduction

The high involvement rate of alcohol use in fatal injuries is well documented.[1–12] For example, blood alcohol has been found in about 40% to 60% of motor vehicle crash victims,[1,4,5] 32% to 46% of homicide victims,[4,6] 20% to 50% of suicides,[4,7] 25% to 30% of drowning victims,[8,9] and 40% to 64% of fire and burn fatalities.[4,10,11] Although alcohol use is generally considered one of the most important risk factors for almost all injuries, the causal role of alcohol use in injuries, except in motor vehicle injury, is largely unproven.[9,12] Several studies have suggested that only heavy drinking is associated with a significantly increased risk of injury.[13,14] One of the studies found a strong association with the number of drinks per occasion but little association with average weekly consumption.[13]

Although alcohol use is directly related to some diseases (e.g., cirrhosis of the liver), the effects of alcohol use on total mortality and coronary heart disease appear to form a U-shaped curve, with the lowest mortality among moderate drinkers and increased mortality among abstainers and heavy drinkers. The public health implications of these findings, however, are not clear because of concerns over other harmful effects of increased alcohol consumption.[15–17]

Most studies relating alcohol to injury or disease have been conducted in limited, select population groups. The purposes of this study were to examine the drinking patterns of a nationally representative sample of US adults aged 25 through 64 years who died in 1986 and the relationship between drinking behavior and the underlying cause of death.

Methods

Data for this study were from the 1986 National Mortality Followback Survey conducted by the National Center for Health Statistics. This survey was designed to collect information on characteristics of decedents that may have affected mortality but were not available on the death certificates. About 1% (18 733) of all adults aged 25 years or older who died in 1986 in the United States, excluding the state of Oregon, were included in this survey, which was based on a stratified random sampling strategy. Information on decedents was obtained by mail questionnaire, telephone, or personal interview from the informants named on the death certificates. The overall response rate was 89%.[18]

Three questions from the survey were used for this analysis. First, a screening question was asked to separate drinkers from nondrinkers: "In his/her entire adult life, did he/she have at least 12 drinks of any kind of alcoholic beverage, such as beer, wine, or liquor?" On frequency of drinking: "On the average, during adult life, how often did he/she drink any alcoholic beverages, such as beer, wine, or liquor?" Finally, consumption per occasion: "On the days that he/she drank, how many drinks did he/she have on the average, per day?" These questions were especially designed to collect information about usual prac-

The authors are with the Injury Prevention Center, Johns Hopkins University School of Hygiene and Public Health, Baltimore, Md.

Requests for reprints should be sent to Guohua Li, MD, DrPH, Injury Prevention Center, Johns Hopkins University School of Hygiene and Public Health, 624 N Broadway, Baltimore, MD 21205.

This paper was accepted December 14, 1993.

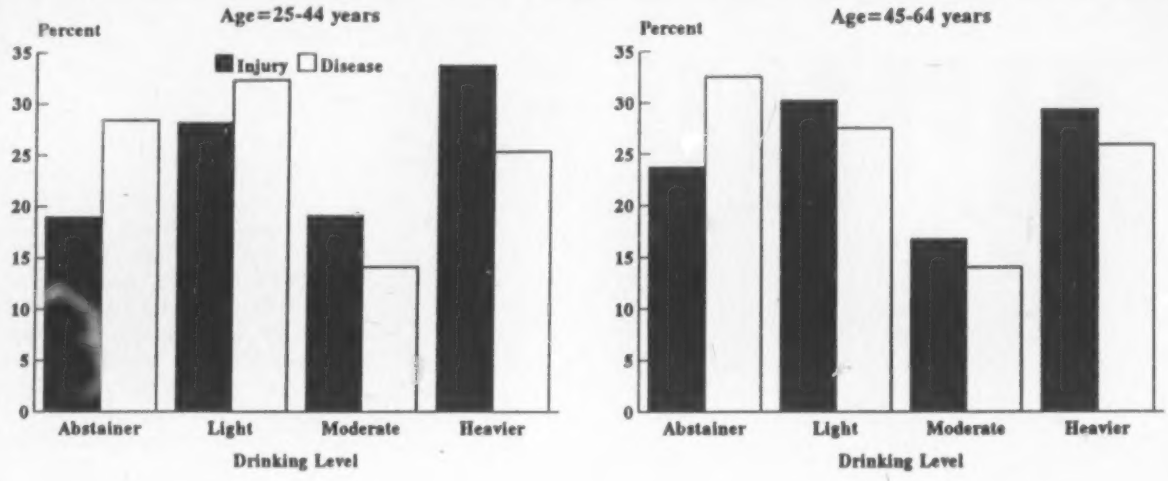


FIGURE 1—Percentage distribution of drinking levels, by decedents' age and cause of death, National Mortality Followback Survey, 1986.

tice rather than any recent change in drinking because of illness.

Excluded from this study were 2135 persons lost to followback, 1872 persons for whom information on any of the three questions was missing, 8173 persons aged 65 or older, and 198 persons whose underlying cause of death was chronic liver disease and cirrhosis (International Classification of Diseases, Ninth Revision[19] [ICD-9] code 571). Persons who died at age 65 or older were excluded because of the difficulty of selecting a single underlying cause of death and the substantial underreporting of injury as the underlying cause of death in the elderly.[20,21] Persons whose underlying cause of death was chronic liver disease or cirrhosis were excluded because alcohol use is very highly correlated with these diagnoses.[22] Thus, the sample size available for data analysis was 6355. Persons who were excluded from this study because their proxies failed to respond or because information on their alcohol use was missing were similar to the sample in age, race, and education but were more likely to be male and unmarried.

Drinking behavior was measured by usual frequency of drinking, usual number of drinks per occasion, and drinking level. Drinking level was determined by the well-established quantity-frequency method, using standard categories[22]: an abstainer was defined as a person who drank fewer than 12 drinks per year; a light drinker, one who drank up to 3 drinks per week; a moderate drinker, one who drank 4 to 13 drinks per week; and a heavier drinker, one who drank 14 drinks or more per week. A standard drink was defined as about 15 g of ethanol.

The underlying cause of death was divided into disease (ICD-9 codes 001–799) and injury (ICD-9 codes E800–E999). Multivariate logistic regression was used to evaluate the effects of drinking behavior on the cause of death while adjusting for demographic characteristics.

Results

The study sample included 5075 persons who died of disease and 1280 who died of injury. Those who died of injury were significantly younger and were more likely to be male, White, Native American, divorced, separated, or never married than persons who died from disease ($P < .01$).

Frequency of Drinking

About 17% of all decedents never drank, and 51% drank at least once a week. Those who died of disease were more likely to be nondrinkers than were those who died of injury (18% vs 12%, $P < .01$). About 40% of those who died of disease either never drank or were infrequent drinkers (less than once per month), compared with 26% for those who died of injury. In contrast, those who died of injury were more likely to be daily drinkers or frequent drinkers (at least once a week) than were those who died of disease, in both sexes and all age groups.

The frequency of drinking differed greatly by sex. Among men, 11% never drank and 62% drank at least once a week, compared with 26% and 33%, respectively, among women. White decedents were most likely to be infrequent drinkers (21%) and Native Americans had the highest rate of daily drinking (23%). Divorced and separated decedents were the most likely of all marital status groups to drink daily (25% for both groups, compared with 15% for married decedents; $P < .01$).

Number of Drinks per Occasion

Overall, 41% of the study sample usually consumed three or more drinks per occasion, and 22% usually consumed five or more drinks per occasion. The prevalence of binge drinking (i.e., five drinks or more per occasion) for those who died of injury was 1.4 times that of persons who died of disease (29% vs 21%, $P < .01$).

The prevalence of binge drinking varied little by age. Male decedents were 2.4 times as likely as female decedents to be binge drinkers (29% vs 12%, $P < .01$). Native American decedents drank much more heavily than members of other races; 50% of the Native American decedents usually consumed five or more drinks per occasion, compared with 21%

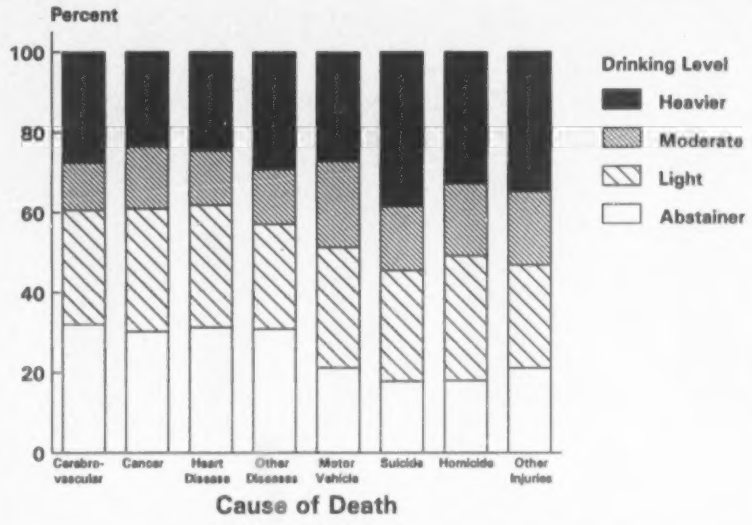


FIGURE 2—Percentage distribution of drinking levels among major categories of disease and injury, decedents aged 25 through 64 years, National Mortality Followback Survey, 1986.

TABLE 1—Percentage Distribution of Drinking Levels of Decedents, by Demographic Characteristics: National Mortality Followback Survey, 1986

	Abstainer	Light Drinker	Moderate Drinker	Heavier Drinker
Age at death, y				
25–34	23.2	31.2	18.1	27.5
35–44	26.5	30.6	14.2	28.7
45–54	30.9	28.6	15.0	25.5
55–64	32.9	26.9	13.4	26.8
Sex				
Male	19.7	27.1	17.9	35.3
Female	42.8	32.5	10.4	14.3
Race				
White	28.2	31.0	14.9	25.9
Black	29.9	25.3	15.5	29.3
Native American	18.7	22.2	18.2	40.9
Other	46.4	24.6	5.8	23.2
Education, y				
<12	27.4	24.5	15.7	32.3
12	29.9	31.0	13.5	25.6
13–15	26.6	34.3	16.0	23.1
≥16	27.7	34.5	16.3	21.5
Marital status				
Married	29.3	33.1	15.3	22.3
Widowed	39.2	24.8	9.4	26.6
Divorced	21.1	23.9	15.4	39.6
Separated	24.2	22.8	13.6	39.4
Never married	30.7	26.0	15.8	27.5

Note. All percentages are row percentages.

of the White decedents and 23% of the Black decedents ($P < .01$). Decedents with less than 12 years of education were 2.7 times as likely as decedents with more than 15 years of education to be binge drinkers. The highest prevalence rate of binge drinking was among decedents who were divorced or separated (34%), followed by those who had never been married (22%), those who were married (18%), and those who were widowed (17%).

Drinking Level

The prevalence of heavier drinking in the study sample was 27%. About a third of those who died of injury (33%) were heavier drinkers, compared with a quarter of those who died of disease ($P < .01$). The association of drinking behavior with cause of death was more striking among young adults than among older adult decedents (Figure 1).

Drinking patterns were similar among the major categories of disease and among the major categories of injury (Figure 2). As found earlier, drinking levels were higher in the injury groups.

High rates of heavier drinking were found among men, Native Americans, persons with less than 12 years of education, and the divorced or separated. The prevalence of heavier drinking was fairly stable over age (Table 1).

Logistic Regression Models

Usual frequency of drinking, number of drinks per occasion, and drinking level were consistently associated with an elevated risk of injury as the underlying cause of death, and each appeared to have a dose–response effect (Table 2). However, the adjusted odds ratios were much smaller than the corresponding crude odds ratios. The effects of demographic variables were similar in the three models. The most important determinant of cause of death was age: the younger the decedent, the more likely it was that the cause of death was injury.

Discussion

This study revealed that 58% of decedents aged 25 through 64 years were either abstainers or light drinkers, and more than a quarter (27%) were heavier drinkers. About 17% of the decedents drank daily, and 22% of all decedents usually consumed five drinks or more per occasion. Drinking patterns across demographic characteristics were consistent with the findings from the 1984 National Alcohol Survey.[23] In both the general population and the decedents, it appeared that the young, Native Americans, and divorced or separated persons were more likely to drink frequently and heavily. Unfortunately, the National Mortality

Followback Survey did not collect information on adults younger than 25 years, a high-risk group for alcohol involvement. Because demographic characteristics were also related to the cause of death, they were strong confounding variables when we examined the effect of drinking behavior on the cause of death. This was illustrated by the discrepancy between the crude odds ratios and the adjusted odds ratios. When adjustments were made for these demographic variables, the association of drinking behavior with the underlying cause of death was weakened, but it was still statistically significant.

The information on drinking behavior was based on responses from proxies, who were primarily the decedents' next of kin. Many studies have assessed the agreement between self- and proxy-reported information on drinking and drinking problems.[24,25] No consensus has emerged about the validity of proxy-reported drinking problems. However, for information on "usual" frequency and quantity of alcohol use, that is, the type of information that this study used, the concordance between self-reports and proxy reports is quite high (60%–70%).[24] Even though informants were explicitly instructed that information on the decedent was about usual practice (i.e., before the decedent became acutely ill), drinking behavior could be more likely to be underreported for persons who died of disease than for those who died of injury because patients with chronic conditions tend to reduce their drinking. Such underreporting would lead to somewhat overestimated odds ratios.

It is noteworthy that we could not determine the association of drinking behavior with the risk of fatal injury or other causes of death because our comparison group was not a representative sample of the general population. Previous studies of alcohol use and the risk of nonfatal injury in which diseased persons were used as a comparison group reported contradictory results. Two cohort studies reported that alcohol use was not associated with a significantly increased risk of fatal injury.[26,27] Other studies have found that an increased risk of fatal injury was associated with the quantity of alcohol consumed (number of drinks per occasion),[13,28] recent history of being drunk,[29] and the frequency of hangover.[30]

Drinking level based on the quantity-frequency method, although typically used to study the adverse health consequences of alcohol use, may be less reliable for estimating injury risk. Regular consumption of 2 or 3 drinks per day places a person in the heavier drinking category, yet the person may never achieve a blood alcohol concentration that would increase the risk of injury. On the other hand, an infrequent binge drinker could be categorized as a light drinker according to the quantity-frequency method, but on occasion such a person would achieve a high blood alcohol concentration and be at high risk of injury. Such a relationship was suggested by Anda et al.,[13] who found the risk of fatal injury to be related to quantity of alcohol consumed per occasion but not total consumption over a 1-month period.

Drinking and injury could interact in two ways. First, the increased risk of injury may result from drinking behavior that exposes the subject to more hazardous circumstances, for example, drinking in bars where the risk of assault is high or driving at excessive speed. Second, the increased risk of injury may be caused by the direct biological effect of alcohol, that is, a reduction in the person's ability to perceive and respond to hazards, such as has been well demonstrated for motor vehicle–related injury.[3] The direct biological effect can be further divided into two components: the direct effect of increased injury risk with increasing blood alcohol concentration, and the hangover or other chronic effects of alcohol, which may increase injury risk even when the blood alcohol concentration is zero. While our study could not separate out these effects, it does provide valuable information about the degree to which usual patterns of alcohol consumption affect injury risk. Most other studies have measured only direct alcohol involvement at the time of injury, and they have assumed if the blood alcohol concentration is not elevated that alcohol is not a factor. A recent study of problem drinking among trauma center patients, however, found the prevalence of problem drinking to be high even in persons who were sober at the time they were injured.[31] It is likely that the different effects of alcohol can interact with each other and affect injury risk in a variety of ways, as suggested by our finding that all three measures of drinking behavior were associated with the likelihood of injury's being the underlying cause of death. □

TABLE 2—Crude and Adjusted Odds Ratios (ORs) and 95% Confidence Intervals (CIs) for Injury as Cause of Death: National Mortality Followback Survey, 1986

	Crude		Adjusted[a]	
	OR	95% CI	OR	95% CI
Frequency of drinking				
Never	1.0	...	1.0	...
< 1 time/mo	1.1	0.9, 1.4	1.0	0.8, 1.3
1–3 times/mo	1.5	1.2, 2.0	1.1	0.9, 1.5
1–2 times/wk	2.0	1.6, 2.5	1.3	1.0, 1.7
3–6 times/wk	2.1	1.7, 2.7	1.4	1.0, 1.8
Every day	1.8	1.5, 2.3	1.4	1.1, 1.8
Number of drinks per occasion				
Never	1.0	...	1.0	...
1	1.1	0.9, 1.4	1.0	0.8, 1.4
2	1.7	1.3, 2.1	1.2	1.0, 1.6
3–4	1.9	1.5, 2.4	1.2	0.9, 1.6
5–6	2.3	1.8, 2.9	1.5	1.1, 2.0
7–11	2.1	1.5, 2.8	1.5	1.1, 2.2
≥ 12	2.1	1.6, 2.8	1.4	1.0, 2.0
Drinking level				
Abstainer	1.0	...	1.0	...
Light	1.5	1.3, 1.8	1.2	1.0, 1.4
Moderate	2.1	1.7, 2.5	1.4	1.1, 1.7
Heavier	2.0	1.7, 2.4	1.4	1.1, 1.7

[a]Adjusted for age, sex, race, education level, and marital status.

Acknowledgments

This research was supported by a grant from the Centers for Disease Control and Prevention (R49/CCR302486) to the Johns Hopkins Injury Prevention Center and by a First Award from the National Institute of Alcohol Abuse and Alcoholism (R29AA07700) to Dr Smith.

References

1. Baker SP, O'Neill B, Ginsburg MJ, Li G. *The Injury Fact Book.* 2nd ed. New York, NY: Oxford University Press; 1992.
2. Centers for Disease Control. Alcohol as a risk factor for injuries—United States. *MMWR.* 1983;32:64.
3. Council on Scientific Affairs. Alcohol and the driver. *JAMA.* 1986;255:522–527.
4. Abel E, Zeidenberg P, Regan S, et al. Alcohol and violent death—Erie County, New York, 1973–1983. *MMWR.* 1984;33: 226–227.
5. Vegega ME, Klein TM. Trends in alcohol-related traffic fatalities, by sex—United States, 1982–1990. *MMWR.* 1992;41:189.
6. Goodman RA, Mercy JA, Loya F, et al. Alcohol use and interpersonal violence: alcohol detected in homicide victims. *Am J Public Health.* 1986;76:144–149.
7. Hlady WG, Middaugh JP. Suicides in Alaska: firearms and alcohol. *Am J Public Health.* 1988;78:179–180.
8. Dietz PE, Baker SP. Drowning: epidemiology and prevention. *Am J Public Health.* 1974;64:303–312.
9. Howland J, Hingson R. Alcohol as a risk factor for drownings: a review of the literature (1950–1985). *Accid Anal Prev.* 1988;20:19–25.
10. Runyan CW, Bangdiwala SI, Linzer MA, Sacks JJ, Butts J. Risk factors for fatal residential fires. *N Engl J Med.* 1992;327: 859–863.
11. Mierley MC, Baker SP. Fatal house fires in an urban population. *JAMA.* 1983;249: 1466–1468.
12. Smith GS, Kraus JF. Alcohol and residential, recreational, and occupational injuries: a review of the epidemiologic evidence. *Ann Rev Public Health.* 1988;9:99–121.
13. Anda RF, Williamson DF, Remington PL. Alcohol and fatal injuries among US adults. *JAMA.* 1988;260:2529–2532.
14. Andreasson S, Allebeck P, Romelsjo A. Alcohol and mortality among young men: longitudinal study of Swedish conscripts. *BMJ.* 1988;296:1021–1025.
15. Shaper AG. Alcohol, the heart, and health. *Am J Public Health.* 1993;83:799–801. Editorial.
16. Stampfer MJ, Rimm EB, Walsh DC. Alcohol, the heart, and public policy. *Am J Public Health.* 1993;83:801–804. Commentary.
17. Peele S. The conflict between public health goals and the temperance mentality. *Am J Public Health.* 1993;83:805–810.
18. Seeman I. National Mortality Followback Survey: 1986 summary, United States. *Vital Health Stat [20].* 1992; no. 19. DHHS publication PHS 92–1856.
19. *Manual of the International Classification of Diseases, Injuries and Cause of Death,* Ninth Revision. Geneva, Switzerland: World Health Organization; 1977.
20. Israel RA, Rosenberg HM, Curtin LR. Analytical potential for multiple cause-of-death data. *Am J Epidemiol.* 1986;124:161–179.
21. Fife D. Injuries and deaths among elderly persons. *Am J Epidemiol.* 1987;126:936–941.
22. Stinson FS, Parrish K, Noble J. Alcohol consumption in a 1986 sample of deaths. *Alcohol Health Res World.* 1990;14:154–162.
23. Clark WB, Hilton ME. *Alcohol in America: Drinking Practices and Problems.* Albany, NY: State University of New York Press; 1991.
24. Lemmens P. *Measurement and Distribution of Alcohol Consumption.* Limburg, the Netherlands: University of Limburg; 1991. Dissertation.
25. Midanik L. Validity of self-reported alcohol use: a literature review and assessment. *Br J Addict.* 1988;83:1019–1029.
26. Gordon T, Doyle JT. Drinking and mortality: the Albany Study. *Am J Epidemiol.* 1987;125:263–270.
27. Cullen K, Stenhouse NS, Wearne KL. Alcohol and mortality in the Busselton Study. *Int J Epidemiol.* 1982;11:67–70.
28. Hingson RW, Lederman RI, Walsh DC. Employee drinking patterns and accidental injury: a study of four New England states. *J Stud Alcohol.* 1985;46:298–303.
29. Kozarevic D, Vojvodic N, Gordon T, et al. Drinking habits and death: the Yugoslavia cardiovascular disease study. *Int J Epidemiol.* 1983;12:145–150.
30. Poikolainen K. Inebriation and mortality. *Int J Epidemiol.* 1983;12:151–155.
31. Soderstrom CA, Dischinger PC, Smith GS, McDuff DR, Hebel JR, Gorelick DA. Psychoactive substance dependence among trauma center patients. *JAMA.* 1992;267: 2756–2759.

Maternal Smoking during Pregnancy and Smoking by Adolescent Daughters

Denise B. Kandel, PhD, Ping Wu, PhD, and Mark Davies, MPH

ABSTRACT

Objectives. Since cigarette smoking in adolescence represents a crucial entry point in the progression to illicit drugs, risk factors for adolescent smoking have public health implications. The influence of mothers on children's smoking appears to be greater than that of fathers. To explain the selective influence of mothers, we examined the consequences of maternal smoking during pregnancy in two longitudinal samples.

Methods. Analyses were conducted on follow-up interview data from two dyadic samples of mothers and firstborn adolescents for whom data on maternal smoking during and after pregnancy were available (192 mother–child pairs originating from New York State and 797 dyads from a national sample).

Results. In both samples, maternal smoking during pregnancy, when postnatal smoking was controlled, selectively increased the probability that female children would smoke and would persist in smoking (adjusted odds ratios of about 4).

Conclusions. The findings suggest that nicotine or other substances released by maternal smoking can affect the fetus, perhaps through the nicotinic input to the dopaminergic motivational system, so as to predispose the brain in a critical period of its development to the subsequent addictive influence of nicotine consumed more than a decade later in life. (*Am J Public Health.* 1994;84:1407–1413)

Introduction

Adolescents who experiment with drugs, and those who ultimately become heavy users, almost invariably first use cigarettes or alcohol.[1–3] Furthermore, the earlier an adolescent begins to experiment with cigarettes and alcohol, the greater the severity and persistence of his or her subsequent involvement with illicit drugs.[3] Thus, it is important to understand the factors that predict early initiation of drinking and cigarette smoking. Among these factors, smoking by parents seems particularly important.

To examine intergenerational effects on drug behavior, we followed a cohort of New York State adolescents for 19 years. In a separate analysis based on mother-father-child triads, we found that maternal smoking was related in a dose-dependent manner to smoking among both boys and girls.[4] The association with paternal smoking was weaker for both sexes. Differences between mothers and fathers were not explained by social influences. Here we report further analyses to identify the sources of the maternal effect in two longitudinal samples, one originally from New York State, the second a national sample. We focus on one factor that differentiates mothers from fathers, namely the mother's smoking behavior during her pregnancy.

Methods

The New York State Follow-Up Cohort

The New York State (NYS) Follow-Up cohort is from a representative sample of adolescents who were enrolled in grades 10 and 11 in NYS public high schools in 1971/72 and were followed for 19 years. The target population for the adult follow-up (n = 1651) was drawn from the enrollment list and included students absent from school at the initial 1971 survey to ensure the representativeness of the sample. Respondents were reinterviewed in 1980, 1984, and 1990, at the average ages of 24–25, 28–29, and 34–35 years; in 1990, 540 men and 620 women (72% of the original target group, excluding those deceased) were reinterviewed. Personal interviews were also conducted with the subjects' two oldest children aged 9–17 years and, when there was a child aged 6–17 years, with the subjects' spouses or partners. The completion rate for the first-born child was 90.5%. Informed consent was obtained from mothers for their own and their children's participation, and separately from the children. The subjects were 192 mothers (mean age = 34.5 years, SD = 0.8) and their first-born children aged 9–17 years (mean age = 12.9, SD = 2.4).

The structured personal interviews took on average 1½ hours to complete for adults and 1 hour for children. Information about the use of 12 classes of drugs within the last 12 months and detailed retrospective monthly drug use histories since the prior survey were obtained at each follow-up. Mothers' use of cigarettes and other drugs during pregnancy with

Denise B. Kandel is with the Department of Psychiatry and the School of Public Health, Columbia University, and the New York State Psychiatric Institute, New York, NY. Ping Wu is with the Department of Psychiatry, Columbia University. Mark Davies is with the New York State Psychiatric Institute.

Requests for reprints should be sent to Denise B. Kandel, PhD, Department of Psychiatry, Columbia University, 722 W 168th St, Box 20, New York, NY 10032.

This paper was accepted April 5, 1994.

Editor's Note. See related editorial by Paneth (p 1372) in this issue.

TABLE 1—Child Smoking Ever and during the Last Year in Two Cohorts, by Maternal Smoking during Pregnancy and Sex of Child

Cohort	Percentage of Children Who Smoked							
		Mother Smoked during Pregnancy						
	Mother Did Not Smoke during Pregnancy	All Smokers	Amount Smoked per Day		Joint χ^2 (df = 1)[a]	P	Joint χ^2 (df = 2)[b]	P
			<1 Pack	≥1 Pack				
A. NYS cohort								
Boys (n)	(71)	(38)	(23)	(15)				
Ever	24.2	26.8	20.4	36.3	0.1	NS	1.3	NS
Last year	14.8	20.2	14.9	28.1	0.5	NS	1.6	NS
Girls (n)	(43)	(41)	(23)	(18)				
Ever	13.2	39.2	46.8	29.4[c]	7.3	<.01	8.9	<.05
Last year	4.3	26.4	24.0	29.4[d]	8.0	<.01	8.2	<.05
B. NLSY cohort								
Boys (n)	(251)	(148)	(115)	(33)				
Ever	17.7	27.9	28.2	26.9[e]	5.8	<.05	5.8	<.10
Last 3 months	3.2	7.7	7.5	8.6	4.0	<.05	4.1	NS
Girls (n)	(249)	(148)	(118)	(30)				
Ever	19.3	20.0	18.2	26.7	0.02	NS	1.1	NS
Last 3 months	5.2	14.4	13.6	17.5[f]	10.0	<.01	10.5	<.01

Note. NYS = New York State; NLSY = National Longitudinal Survey of the Work Experience of Youth; NS = not significant.
[a]Test of independence between child's smoking and any maternal smoking during pregnancy.
[b]Test of independence between child's smoking and amount smoked by mother during pregnancy.
[c]No smoking vs <1 pack, P < .01.
[d]No smoking vs <1 pack, P < .05; no smoking vs ≥1 pack, P < .01.
[e]No smoking vs <1 pack, P < .05.
[f]No smoking vs < 1 pack, P <.01; no smoking vs ≥1 pack, P < .01.

the selected child and amount smoked were ascertained.

The National Longitudinal Survey of the Work Experience of Youth Cohort

The analyses were replicated on a second group of subjects obtained from the National Longitudinal Survey of the Work Experience of Youth (NLSY). This multistage stratified national probability sample, representative of youths born from 1957 through 1963 in the coterminous United States, was drawn in 1979 and interviewed annually thereafter. The 1979 interview completion rate was 90%; retention rates have consistently been over 90%. In 1990, 4510 women (91% of the target group) were interviewed, including 3088 mothers; 5803 children (90% of the target group) born to these mothers were assessed. Interviews were completed with 957 mothers and their oldest children aged 10–18 years.

Analyses were based on 797 of these dyads with complete data on child smoking and maternal smoking during and after pregnancy (mothers' mean age = 30.7 years, SD = 1.9; children's mean age = 12.5 years, SD = 1.9). Prenatal exposure to maternal tobacco and alcohol use was first ascertained in 1983 for the youngest child and in 1986 for all births not asked about previously (62.6% of the children). Questions about current maternal drug use, including cigarette smoking, were asked only in 1984. In 1990, children aged 10 years and older completed a self-administered questionnaire inquiring about cigarette smoking. Informed consent was obtained from mothers for their own and their children's participation.[5]

The NLSY data differ from the NYS data in several respects: (1) mothers were specifically asked whether they had smoked during the 12 months preceding the birth of their children and how much they smoked during the pregnancy; (2) the most recent measure of maternal smoking was obtained in 1984, 6 years prior to the children's assessment; (3) monthly drug histories were unavailable; and (4) children's current smoking was ascertained for the past 3 months, not for the last year.

Statistical Analysis

Pearson's chi-square test was used to test for homogeneity of rates in contingency tables. When this joint test was significant, pairwise rates in the 2 × 2 subtables were compared by using Pearson's chi-square test. Maximum likelihood estimation and hypothesis testing methods were used in the multivariable logistic regressions of binary indicators of the child's smoking behaviors on maternal age, education, and smoking behaviors.[6] The large sample estimates of the standard errors were used to construct confidence intervals. Odds ratios were pooled across samples and tested for significance with the Mantel-Haenszel method.[7,8] The statistical program SPSS (SPSS Inc, Chicago, Ill) was used for weighted calculations.

Results

NYS Analyses

Of the mothers, 40.7% smoked during pregnancy; 28.7% smoked throughout the pregnancy. Twenty-five percent of the male offspring and 25.8% of the female offspring ever smoked; 16.7% and 15.0%, respectively, smoked within the last 12 months. There is a significant association between maternal smoking during pregnancy and the child's smoking 13 years later (Table 1). This association is stron-

ger for the child's smoking during the last year than for ever smoking, and it is stronger for daughters than for sons. We controlled for maternal use of alcohol and marijuana during pregnancy and found that the smoking effect was not altered (data not shown). Only one woman reported using illicit drugs other than marijuana during her pregnancy.

The proportion of current smokers among ever smokers can be interpreted as a measure of persistence of smoking. Persistence of smoking among offspring is more highly related to maternal smoking during pregnancy for daughters than for sons (Table 2, A.).

Of mothers who smoked during their pregnancies, 77.7% were still smoking 13 years later. Since the effect of smoking during pregnancy could be confounded with the effect of current smoking, we carried out two analyses to examine the joint impact of maternal smoking behavior during and after pregnancy.

TABLE 2—Persistence of Child Smoking in Two Cohorts, by Maternal Smoking during Pregnancy and Sex of Child

Cohort	Percentage of Children Who Continued to Smoke[a]: Mother Did Not Smoke during Pregnancy %	(n)	Mother Smoked during Pregnancy %	(n)	Joint χ^2 ($df = 1$)	P
A. NYS cohort						
Boys	61.4	(17)	75.3	(10)	0.6	NS
Girls	32.3	(6)	67.3	(16)	2.1	NS
B. NLSY cohort						
Boys	17.4	(46)	27.6	(41)	1.3	NS
Girls	26.9	(48)	72.2	(30)	15.2	<.001

Note. NYS = New York State; NLSY = National Longitudinal Survey of the Work Experience of Youth; NS = not significant.

[a]Percentage who continued to smoke (i.e., during the last 12 months in the New York State cohort and during the last 3 months in the National Longitudinal Survey of the Work Experience of Youth) among those who ever smoked.

In the first step, five groups of mothers were identified, depending upon whether they smoked during the pregnancy and within the year preceding the interview (Table 3, A.). The child's exposure to postnatal maternal smoking varied greatly across the groups. On average, mothers in group 2 had last smoked 14.3 years ago, mothers in group 4 had last smoked 6.5 years ago, and mothers in groups 3 and 5 were currently smoking. Total exposure, calculated by dividing the total number of months the mother smoked after childbirth by the child's age in months, ranged from 4% when mothers smoked neither during pregnancy nor within the last year (group 2) to 38% when mothers smoked during pregnancy but not within the last year (group 4), to 63% when mothers did not smoke during pregnancy but smoked within the last year (group 3), and to 91% when mothers smoked in both periods (group 5). The effect of maternal smoking during pregnancy on daughters is apparent across the groups. Rates of smoking among daughters were higher when mothers smoked during pregnancy but not within the last year (group 4) than when mothers did not smoke during pregnancy but did smoke within the last year (group 3). Yet the children of group 3 mothers were exposed to maternal smoking for twice as long in their lives, and much more recently, than the children of the group 4 mothers. Correlatively, the same proportions of daughters were currently smoking if their mothers smoked during pregnancy, irrespective of whether the mothers smoked within the last year. However, daughters of the currently smoking mothers (group 5) were exposed to postnatal maternal smoking for three times as long as daughters of the mothers who had not smoked within the last year (group 4); the latter mothers had last smoked more than 6 years ago. By contrast, recent smoking by sons appears to be more influenced by maternal smoking within the last year. The relationship of daughters' smoking rates to maternal smoking during pregnancy and the lack of relationships to child lifetime exposure and current maternal smoking further support the existence of a prenatal effect of maternal smoking.

In the second step, we estimated logistic regressions predicting the child's smoking—ever and within the year preceding the interview—in which we distinguished the amounts mothers smoked, both during pregnancy and currently (Table 4). (The second model shown did not differentiate maternal smoking during pregnancy according to quantities smoked.) The strongest positive effect of maternal smoking appears when the mother smoked during her pregnancy. This effect is evident only among daughters and is quite marked, regardless of the amount the mother smoked. By contrast, the mother's current smoking is nonsignificant but has a greater effect on boys than on girls.

A Replication in the NLSY Sample

To test the NYS results by replication, we carried out a parallel analysis in the NLSY sample. Of the mothers, 37.1% smoked during pregnancy; 21.5% of the boys and 19.6% of the girls ever smoked and 4.9% and 8.6%, respectively, smoked within the last 3 months. The child's smoking, especially current and persistent smoking (the proportion who smoked in the last 3 months among those who ever smoked), is related to maternal smoking during pregnancy (Tables 1 and 2, B.). As in the NYS sample, the effect was stronger on daughters than on sons.

Also as in the NYS sample, smoking during pregnancy was highly related to smoking afterward: 89% of mothers who had smoked during pregnancy were smoking at the time of the 1984 interview (an average of 6 to 7 years after pregnancy). The joint classification of maternal smoking during and after pregnancy reveals the stronger impact of smoking during pregnancy than afterward and on daughters than on sons (Table 3, B.).

To identify the unique effect of prenatal maternal smoking, we estimated the same logistic regressions for the NLSY sample as for the NYS sample (Table 5). Again, maternal smoking during pregnancy was associated with current and persistent smoking by daughters but not sons. Contrary to the results of the NYS analysis, there was no increased risk of the child's ever smoking due to maternal smoking during pregnancy, and the significant effect on daughters' current smoking was found among mothers who reported smoking less than a pack of cigarettes a day during pregnancy, rather than a pack or more. Heavy maternal smoking 6 years prior to the child's

TABLE 3—Current Child Smoking in Two Cohorts, by Maternal Smoking History during and after Pregnancy and Sex of Child

	Percentage of Children Who Smoked											
	(1) Mother Never Smoked		(2) Mother Smoked but Not during Pregnancy/ Not Last Year[a]		(3) Mother Did Not Smoke during Pregnancy/ Smoked Last Year[a]		(4) Mother Smoked during Pregnancy/ Not Last Year[a]		(5) Mother Smoked during Pregnancy and Last Year[a]		Joint χ^2	
Cohort	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	($df = 4$)	P
A. NYS (smoked last year)												
Boys	15.2	(21)	10.1	(38)	29.5	(12)	11.3	(11)	23.8	(27)	3.9	NS
Girls	0.0	(13)	6.4	(20)	5.7	(9)	26.9	(6)	26.3[b]	(34)	8.3	<.10
B. NLSY (smoked last 3 months)												
Boys	4.8	(60)	2.5	(96)	3.0	(95)	0.0	(22)	9.1[c]	(126)	7.8	<.10
Girls	2.0	(54)	4.4	(98)	7.8	(97)	27.2	(11)	13.4[d]	(137)	14.5	<.01

Note. NS = not significant.
[a]"Last year" refers to the year preceding the current maternal and child interviews in the New York State (NYS) cohort, and the year preceding the maternal interview 6 years before the current child interview in the National Longitudinal Survey of the Work Experience of Youth (NLSY) cohort.
[b]Group 1 vs group 4, $P < .05$; group 1 vs group 5, $P < .05$; group 2 vs group 5, $P < .10$.
[c]Group 2 vs group 5, $P < .05$; group 3 vs group 5, $P < .10$.
[d]Group 1 vs group 4, $P < .001$; group 1 vs group 5, $P < .05$; group 2 vs group 4, $P < .01$; group 2 vs group 5, $P < .05$; group 3 vs group 4, $P < .05$.

TABLE 4—Adjusted Odds Ratios (AORs) and Confidence Intervals (CIs) from Logistic Regressions Predicting Child Smoking from Maternal Smoking during and after Pregnancy, by Sex of Child: New York State Cohort

	Child Smoking Ever				Child Smoking Last Year			
	Boys (n = 109)		Girls (n = 83)		Boys (n = 109)		Girls (n = 83)	
Predictors[a]	AOR	95% CI	AOR	95% CI	AOR	95% CI	AOR	95% CI
Model 1								
Child's age[b]	2.3*	1.6, 3.3	1.8*	1.3, 2.6	2.0*	1.4, 2.9	1.9**	1.2, 3.0
Maternal smoking during pregnancy								
Did not smoke (reference group)								
<1 pack per day	0.5	0.1, 2.2	5.2***	0.9, 29.5	0.7	0.1, 3.7	3.0	0.3, 32.1
≥1 pack per day	1.1	0.2, 5.8	4.3	0.5, 40.1	1.5	0.3, 8.1	11.2***	0.7, 179.8
Maternal smoking after pregnancy (last year[c])								
Did not smoke (reference group)								
<1 pack per day	3.1	0.6, 17.4	0.2	0.0, 1.6	2.3	0.4, 13.7	0.6	0.1, 7.9
≥1 pack per day	1.8	0.4, 8.0	0.5	0.1, 3.4	1.7	0.4, 7.9	1.2	0.1, 11.6
Constant	−14.6*		−7.1***		−12.9**		−3.8	
χ^2[d] ($df = 6$)	43.0*		27.6*		25.3**		23.5*	
Model 2[e]								
Maternal smoking during pregnancy (vs not smoking)	0.7	0.2, 2.4	5.0***	0.9, 27.7	1.0	0.3, 3.8	4.0	0.4, 38.1

[a]Mother's education was included in the model. Nonsignificant coefficients not shown.
[b]Odds ratios are calculated for a 1-year change in this continuous variable.
[c]"Last year" refers to the year preceding interview.
[d]Maximum likelihood chi-square for the difference between a model including only a constant and the model including the set of independent variables.
[e]Same variables included as in Model 1, except that the two categories for amount smoked by mother were combined. Other coefficients not shown.
*$P < .001$; **$P < .01$; ***$P < .10$.

interview, when the child was almost 7 years old, was associated with increased odds of sons' ever smoking; surprisingly, maternal smoking of less than a pack a day was associated with decreased odds of daughters' ever smoking.

In both cohorts, the adjusted odds ratios of any maternal smoking during pregnancy, irrespective of quantities smoked, on daughters' current smoking are very similar (4.0 in the NYS sample, 4.1 in the NLSY sample; see model 2,

TABLE 5—Adjusted Odds Ratios (AORs) and Confidence Intervals (CIs) from Logistic Regressions Predicting Child Smoking from Maternal Smoking during and after Pregnancy, by Sex of Child: National Longitudinal Survey of the Work Experience of Youth Cohort

	Child Smoking Ever				Child Smoking Last 3 Months				Persistent Smoking[a]			
	Boys (n = 406)		Girls (n = 381)		Boys (n = 406)		Girls (n = 381)		Boys (n = 77)		Girls (n = 65)	
Predictors[b]	AOR	95% CI	AOR	95% CI	AOR	95% CI	AOR	95% CI	AOR	95% CI	AOR	95% CI
Model 1												
Child's age[c]	1.6*	1.4, 1.8	1.7*	1.5, 2.0	1.8*	1.3, 2.3	2.2*	1.7, 2.9	1.3	0.9, 1.8	1.7**	1.2, 2.5
Maternal smoking during pregnancy												
Did not smoke (reference group)												
< 1 pack per day	1.4	0.8, 2.6	1.1	0.5, 2.3	1.5	0.5, 4.7	4.8**	1.6, 14.7	1.0	0.3, 3.5	5.5***	1.3, 22.2
≥ 1 pack per day	1.0	0.4, 2.7	1.0	0.3, 3.0	2.0	0.4, 9.9	1.9	0.4, 9.9	1.4	0.2, 9.2	2.2	0.3, 18.6
Maternal smoking after pregnancy (last year)[d]												
Did not smoke (reference group)												
< 1 pack per day	0.8	0.4, 1.7	0.4***	0.2, 1.0	1.6	0.4, 5.9	0.2***	0.1, 0.9	2.3	0.5, 9.9	0.6	0.1, 2.7
≥ 1 pack per day	2.8**	1.4, 5.7	1.2	0.6, 2.8	3.1	0.8, 12.6	1.4	0.4, 4.5	2.1	0.4, 10.3	1.9	0.4, 8.2
Constant	−7.1*		−8.9*		−11.3*		−13.2*		−5.4		−7.6**	
χ^2[e] (df = 6)	61.0*		67.0*		25.5*		75.4*		5.3		29.6*	
Model 2[f]												
Maternal smoking during pregnancy (vs not smoking)	1.3	0.7, 2.4	1.1	0.5, 2.2	1.6	0.6, 4.6	4.1***	1.4, 12.1	1.1	0.3, 3.5	4.6***	1.2, 17.2

[a]Smoking in last 3 months among children who ever smoked.
[b]Mother's education was included in the model. Nonsignificant coefficients not shown.
[c]Odds ratios are calculated for a 1-year change in this continuous variable.
[d]Measured 6 years prior to the child's interview.
[e]Maximum likelihood chi-square for the difference between a model including only a constant and the model including the set of independent variables.
[f]Same variables included as in Model 1, except that the two categories for amount smoked by mother were combined. Other coefficients not shown.
*$P < .001$; **$P \leq .01$; ***$P < .05$.

Tables 4 and 5). A Mantel-Haenszel chi-square test of the common underlying association between mother's and daughter's smoking across the two samples is statistically significant ($P < .01$).

To test the hypothesis that the effect of maternal prenatal smoking is stronger for daughters' persistence of smoking than for experimentation with cigarettes, we conducted a logistic regression predicting persistent smoking among adolescents who had ever smoked in the NLSY cohort. Maternal smoking during pregnancy increased the odds of persistent smoking by daughters more than four times, but had no effect among sons (Table 5). The adjusted odds ratios for daughters were statistically significantly different from those for sons. The small sample size precluded the implementation of a sex-specific analysis of persistence in the NYS sample.

Discussion

Although the sample sizes were small, the strength of our findings derives from the test of an a priori hypothesis in one sample and a replication in a second sample. In both, the maternal effect of smoking during pregnancy was stronger on daughters than on sons, and in one sample it was clearly stronger for the child's current smoking and for persistence than for lifetime experimentation.

The impact of maternal smoking during pregnancy on the child's smoking, and especially the differential impact on current smoking compared with ever smoking, is consonant with the existence of a prenatal effect. Such a biological effect should be stronger for persistent and heavy smoking than for simple experimentation; the latter should be more responsive to social influences. The inconsistent findings between the two samples regarding the dose effects parallel findings regarding the effect of prenatal smoking in lowering birthweight.[9] It would be especially important to replicate the analyses in a prospective cohort in which detailed maternal drug use could be assessed during pregnancy.

One mechanism underlying the maternal effect and its greater impact on female than male children might be an intrauterine effect of nicotine on the developing brain of the fetus. Some motivational consequences of reinforcing stimuli, including the craving from drug dependence, are mediated by the mesolimbic and mesocortical components of the dopaminergic system.[10-14] Drug withdrawal leads to decreased activity in components of the dopaminergic system that are clinically manifested in craving. Drug intake restores the rewarding effects

of dopaminergic transmission.[15,16] Nicotine, which crosses the placental barrier and stimulates the action of cholinergic neurons, enhances activity in the dopaminergic system.[13]

Our data raise the possibility that during a critical prenatal period of brain development, nicotine, or other substances released by maternal smoking, may modify the properties of the dopaminergic system of the brain so as to change the threshold of this system (or related systems) to the effects of nicotine at a later time in life and thereby predispose the child to smoke and to persist in smoking. This suggestion is consistent with the notion that nicotine, like other addictive drugs, can alter gene expression and can produce long-lasting functional and structural changes in dopaminergic neurons,[10] an effect that might be particularly profound in the developing brain. The finding that nicotine administered to pregnant mice increases adenylate cyclase activity in the midbrain and brainstem at postpartum and that the levels return to normal within 4 weeks, only to recur in adulthood,[17] lends support to this interpretation.

Why the intergenerational transmission of smoking from mother to child should be stronger for female children than for males is unclear. Perhaps these sex differences reflect the distinctive sexual dimorphism of the brain—including hormonal and structural factors—that emerges during fetal development.[18,19] For example, the release of androgens may protect the male infant against the priming effect of nicotine.

Maternal prenatal exposure to drugs may have two types of consequences for the child: immediately manifest consequences, most clearly evident in crack babies or infants with the fetal alcohol syndrome, and latent consequences that may require more than a decade to become manifest, as may be the case for smoking. The prenatally induced predisposition to smoke might be enhanced further by passive smoking after birth, which would facilitate the child's desire to experiment with cigarettes and his or her potential dependence on tobacco. We could not measure exposure to passive smoking by other household members. We assumed that the effects of passive smoking would be captured by postnatal maternal smoking. Since we showed that postnatal maternal smoking did not retain a unique statistically significant effect and since we had previously shown that current maternal smoking was more important than paternal smoking,[4] we conclude that the prenatal maternal effect is stronger than the effect of subsequent passive smoke. Subtle physiological effects deriving from the child's exposure to prenatal maternal smoking may create in these children a low level of nicotine dependence without their ever having smoked a cigarette. These differences could be reflected in the child's initial response to nicotine and result in tolerance and persistence of smoking. Our data do not exclude a variety of additional social or physiological effects, including a genetic predisposition to smoking, reflected in differential regulation of brain nicotinic receptor numbers and response to nicotine.[15,16] Such a genetic predisposition might be especially strong among women who continue to smoke during their pregnancies despite the well-advertised hazards of doing so.

The present findings are consonant with an increasing body of research that documents latent and deferred effects of fetal exposure to a variety of insults not limited to drug use. Thus, the classic studies on the Dutch famine during World War II have identified adult effects of prenatal food deprivation on obesity and schizophrenia.[20,21] Interestingly, the schizophrenia effects are somewhat stronger among women than among men.

Independent of its mechanism, our evidence indicates that maternal smoking during pregnancy may create a serious risk for smoking dependence in female offspring, a risk that is even greater than that created by the mother's current smoking. Young women have not shown the decline in smoking observed among other adults in the United States.[22] From a public health perspective, it is essential to target these young women for prevention and intervention, since a reduction in their rates of smoking may have unanticipated benefits for the second generation in addition to the direct health benefits for the adult smokers themselves. □

Acknowledgments

Work on this manuscript was partially supported by research grants DA03196 and DA04866 (D. Kandel, principal investigator) and a research scientist award (DA00081) to D. Kandel from the National Institute on Drug Abuse. Partial support for computer costs was provided by Mental Health Clinical Research Center grant MH30906 from the National Institute of Mental Health to the New York State Psychiatric Institute.

The research assistance of Christine Schaffran is gratefully acknowledged, as are the suggestions of E. Kandel. This article benefited also from comments by the reviewers of the Journal.

References

1. Kandel D. Stages in adolescent involvement in drug use. *Science.* 1975;190:912–914.
2. Kandel D, Yamaguchi K, Chen K. Stages of progression in drug involvement from adolescence to adulthood: further evidence for the gateway theory. *J Stud Alcohol.* 1992;53:447–457.
3. Kandel D, Yamaguchi K. From beer to crack: developmental patterns of involvement in drugs. *Am J Public Health.* 1993;83:851–855.
4. Kandel D, Wu P. The contribution of mothers and fathers to the intergenerational transmission of cigarette smoking in adolescence. *J Res Adolesc.* In press.
5. *1990 NLSY Child Magnetic Tape User's Documentation.* Columbus, Ohio: The Ohio State University Center for Human Resource Research; 1992. Introduction, Table 1.6.
6. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons Inc; 1989.
7. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22:719–748.
8. Fleiss J. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley & Sons Inc; 1981.
9. Stein Z, Kline J. Smoking, alcohol and reproduction. *Am J Public Health.* 1983;73:1154–1156.
10. Beitner-Johnston D, Guitart X, Nestler EJ. Neurofilament proteins and the mesolimbic dopamine system: common regulation by chronic morphine and chronic cocaine in the rat ventral tegmental area. *J Neurosci.* 1992;12:2165–2175.
11. Corrigall WA. Understanding brain mechanisms in nicotine reinforcement. *Br J Addict.* 1991;86:507–510.
12. Koob GF. Drugs of abuse: anatomy, pharmacology and function of reward pathways. *TiPS.* 1992;13:177–184.
13. Mereu G, Yoon K-WP, Boi V, et al. Preferential stimulation of ventral tegmental area dopaminergic neurons by nicotine. *Eur J Pharmacol.* 1987;141:395–399.
14. Nestler EJ. Molecular mechanisms of drug addiction. *J Neurosci.* 1992;12:2439–2450.
15. Collins AC. Genetic influences on tobacco use: a review of human and animal studies. *Int J Addict.* 1990–1991;25:35–55.
16. Heath AC, Madden PAF. Genetic influences on smoking behavior. In: Turner JR, Cardon LR, Hewitt JK, eds. *Behavioral Genetic Approaches in Behavioral Medicine.* New York, NY: Plenum; in press: chapter 10.
17. Slotkin TA, McCook EC, Lappi SE, Seidler FJ. Altered development of basal and forskolin-stimulated adenylate cyclase activity in brain regions of rats exposed to

nicotine prenatally. *Dev Brain Res.* 1992;68: 233–239.
18. Kelly DD. Sexual differentiation of the nervous system. In: Kandel ER, Schwartz JH, Jessell TM, eds. *Principles of Neural Sciences.* 3rd ed. New York, NY: Elsevier; 1991:959–973.
19. Tallal P, McEwen BS, eds. Neuroendocrine effects on brain development and cognition. *Psychoneuroendocrinology.* 1991; 16:special issue.
20. Susser ES, Lin SP. Schizophrenia after prenatal exposure to the Dutch Hunger winter of 1944–1945. *Arch Gen Psychiatry.* 1992;49:983–988.
21. Ravelli G, Stein ZA, Susser M. Obesity in young men after famine exposure in utero and early infancy. *N Engl J Med.* 1976;295: 349–353.
22. *Preliminary Estimates from the 1992 National Household Survey on Drug Abuse.* Rockville, Md: Office of Applied Studies, Substance Abuse and Mental Health Services Administration; 1993.

Call for Papers on Worksite Health Programs

Health Education Quarterly, the official journal of the Society for Public Health Education, announces a call for original manuscripts for a theme issue on worksite health programs. The worksite is a common context within which health promotion, disease prevention, and injury prevention programs are conducted. This theme issue is intended to serve as a forum for those interested in occupational and worksite health and to foster a dialogue between those involved in worksite health promotion and those in occupational health and safety.

The journal invites practitioners, researchers, policymakers, and interested others to submit manuscripts that accomplish any of the following:

- Describe the development, implementation, and evaluation of a worksite health promotion, disease prevention, or injury prevention intervention (e.g., descriptions of methods for gaining employee and organizational commitment to and active participation in worksite health programs; innovative needs assessments; methods for maintaining interest in and commitment to a program over time; process and/or outcome evaluations of worksite interventions)
- Investigate factors that affect worksite health programs and the practice of health educators within those programs (e.g., research studies addressing organizational, work group, or employee characteristics related to successful worksite health programs)
- Address important issues in worksite health programs and integrate the worksite health promotion and occupational health and safety perspectives (e.g., literature reviews or conceptual "think pieces" that address such crosscutting issues as rectifying health status goals with organizational profit maximization goals; addressing the possibility that either explicit or subtle coercive influences affect employees' decisions to participate in worksite health programs or to practice safe work behaviors; accommodating multiple stakeholders' divergent points of view that contribute to program planning or policy decisions)
- Describe methods or tools available to individuals who conduct or evaluate worksite health interventions (e.g., use of biomarkers, industrial hygiene methods)

All papers will be subject to peer-review procedures. Authors should consult the "Information for Authors" in *Health Education Quarterly* for style requirements. Manuscripts will be accepted until *February 28, 1995,* and should be submitted to Catherine A. Heaney, PhD, MPH, The Ohio State University, Department of Preventive Medicine, B-215 Starling Loving Hall, 320 W 10th Ave, Columbus, OH 43210; tel (614) 293-3908.

An Evaluation of the Kessner Adequacy of Prenatal Care Index and a Proposed Adequacy of Prenatal Care Utilization Index

Milton Kotelchuck, PhD, MPH

ABSTRACT

Objectives. The assessment of the adequacy of prenatal care utilization is heavily shaped by the way in which utilization is measured. Although it is widely used, the current major index of utilization, the Kessner/Institute of Medicine Index, has not been subjected to systematic examination. This paper provides such an examination.

Methods. Data from the 1980 National Natality Survey are used to disaggregate the components of the Kessner Index for detailed analysis. An alternative two-part index, the Adequacy of Prenatal Care Utilization Index, is proposed that combines independent assessments of the timing of prenatal care initiation and the frequency of visits received after initiation.

Results. The Kessner Index is seriously flawed. It is heavily weighted toward timing of prenatal care initiation, does not distinguish timing of initiation from poor subsequent utilization, inaccurately measures utilization for full- or post-term pregnancies, and lacks sufficient documentation for consistent computer programming.

Conclusions. The Adequacy of Prenatal Care Utilization Index offers a more accurate and comprehensive set of measures of prenatal care utilization than the Kessner Index. (*Am J Public Health.* 1994;84:1414–1420)

Introduction

Accurate assessment of prenatal care utilization is the critical first step in the development of public health programs to improve prenatal care accessibility and ultimately to improve birth outcomes. The assessment of the adequacy of prenatal care utilization is heavily shaped by the way in which utilization is measured.

Currently, there are two widely used measures of adequacy of prenatal care utilization: the trimester of prenatal care initiation and the Kessner/Institute of Medicine (IOM) Adequacy of Prenatal Care Index.[1] Trimester of initiation will not be examined here because it provides no information about prenatal care utilization after initiation; it has been critically examined elsewhere.[2,3] The Kessner Index—the principal adequacy of prenatal care utilization index in use in the United States today—includes information about both the timing of prenatal care initiation and prenatal care visits after initiation. It was published in 1973 as part of an IOM-supported study of infant mortality in New York City.[1] The Kessner Index combines two continuous numeric measures (month prenatal care begins and number of visits, adjusting for length of gestation) and rigidly links them into a very easy to understand index with three levels of adequacy (Adequate, Intermediate, and Inadequate). To be rated Adequate on the Kessner Index, one must start prenatal care in the first trimester and have nine prenatal care visits for a normal-length pregnancy.

Table 1 presents the Kessner Index algorithm as initially published. In this original description, public or private obstetric service was a third factor, but this factor has been dropped by all subsequent researchers because the type of service is not noted on the standard US birth certificate and because researchers disagree with the index's assumption that care from public services can never be adequate.[4–7] Although Kessner et al. called their index the "Adequacy of Prenatal Care Index," their measure indicates nothing about the content or clinical adequacy of prenatal care: it is a utilization index only.

The Kessner Index has been widely adopted for public health research, planning, and resource allocation. However, it appears that the index was not subjected to close scrutiny prior to its widespread adoption. The accuracy of the Kessner Index is critical because any limitations may distort our perceptions about prenatal care adequacy in the United States and may incorrectly influence programmatic efforts to improve prenatal care utilization. The Kessner Index has also been widely used to assess the association between prenatal care and birth outcomes.[4–7] The limited positive associations noted may be more a reflection of the internal characteristics of the Kessner Index than of the true strength of that relationship.

The ability to measure prenatal care utilization after initiation remains underdeveloped. The Kessner Index does not separately isolate utilization after enrollment, nor does any other prenatal care index. Yet a distinction between initiation

The author is with the Department of Maternal and Child Health, School of Public Health, University of North Carolina at Chapel Hill.

Requests for reprints should be sent to Milton Kotelchuck, PhD, MPH, Department of Maternal and Child Health, School of Public Health, University of North Carolina at Chapel Hill, CB# 7400, Chapel Hill, NC 27599-7400.

This paper was accepted September 23, 1993.

Editor's Note. See related editorial by Wise (p 1374) and article by Kotelchuck (p 1486) in this issue.

and utilization once in care may have differential implications for birth outcomes and for prenatal care program practices.

In this paper I examine the characteristics and limitations of the Kessner Adequacy of Prenatal Care Index and propose an alternative, the two-factor Adequacy of Prenatal Care Utilization Index.

Database

The database for this paper is the National Center for Health Statistics' 1980 National Natality Survey, a representative sample of US births in 1980.[8] This survey uses a follow-back methodology involving the collection of data from four sources: maternal retrospective information, physician and hospital records, radiologic records, and birth and death certificates. Data are available on 9941 live births, oversampled (4:1) for low-birthweight infants. Responses were weighted by a poststratification ratio estimate procedure to be representative of the 1980 US birth cohort. Prenatal care information is absent for 15% of the births. All missing data were imputed via a categorical matrix hot-deck methodology. Details of the survey and study population are described elsewhere.[9]

All data used in this paper are derived from the birth certificate data source only. The birth certificate data were chosen because they are readily available to the public health community, they are the principal database for the assessment of a community's prenatal care utilization adequacy, and they are available for all married and unmarried women in the 1980 National Natality Survey.

Kessner Index Assessment

The Kessner Index is a seriously flawed index of adequacy of prenatal care utilization. Four features merit attention.

First, the Kessner Index is principally a measure of the timing of initiation of prenatal care. The Kessner Index algorithm requires that to be rated Adequate, prenatal care must begin in the first trimester; to be rated Intermediate, care must begin in the second trimester; and to be rated Inadequate, care must begin in the third trimester or not at all. The additional factor in the Kessner Index, the number of prenatal care visits, can only lower the rating category. This rarely happens. Table 2 shows that the trimester of care overwhelmingly (for 86.2% of women) predicts the Kessner Index rating. Only 13.8% of women have their ratings reduced owing to insufficient visits.

Second, the Kessner Index does not distinguish inadequacy of care due to late initiation from inadequacy of care due to insufficient number of visits. Table 3 shows that these two implicit subscales have distinctive distributions. Overall, 24.7% of US women would be classified differently if the two were measured separately. More striking, 61.1% of women starting prenatal care in the second trimester (rated Intermediate by the Kessner Index) and 45.6% of women starting prenatal care in the last trimester or not at all (rated Inadequate by the Kessner Index) would be classified differently based on an index with two distinctive factors.

Third, the Kessner Index is unable to adequately characterize prenatal care utilization for normal-gestation and postmature births. For all normal-length pregnancies (more than 36 weeks' gestation), the Kessner Index requires only nine visits for care to be Adequate. Yet up to 36 weeks' gestation, the Kessner Index adjusts the

TABLE 1—Three-Factor Health Services Index Controlled for Gestation and Based on Number of Prenatal Visits, Interval to First Prenatal Visit, and Type of Hospital Service

Medical Care Index	Gestation (Weeks)		Number of Prenatal Visits
Adequate[a]	13 or less	and	1 or more or not stated
	14–17	and	2 or more
	18–21	and	3 or more
	22–25	and	4 or more
	26–29	and	5 or more
	30–31	and	6 or more
	32–33	and	7 or more
	34–35	and	8 or more
	36 or more	and	9 or more
Inadequate[b]	14–21[c]	and	0 or not stated
	22–29	and	1 or less or not stated
	30–31	and	2 or less or not stated
	32–33	and	3 or less or not stated
	34 or more	and	4 or less or not stated
Intermediate	All combinations other than specified above		

[a]In addition to the specific number of visits indicated for adequate care, the interval to the first prenatal visit had to be 13 weeks or less (first trimester), and the delivery must have taken place on a private obstetrical service.
[b]In addition to the specific number of visits indicated for inadequate care, all women who started their prenatal care during the third trimester (28 weeks or later) were considered inadequate.
[c]For this gestation group, care was considered inadequate if the time of the first visit was not stated.
Source. Reprinted with permission from *Infant Death: An Analysis by Maternal Risk and Health Care* (Table 2-3, p 59). Copyright ©1973, National Academy of Sciences. Courtesy of the National Academy Press, Washington, DC.

TABLE 2—Ratings Assigned to Births According to the Kessner Adequacy of Prenatal Care Index, by the Trimester of Initiation of Prenatal Care

Trimester of Initiation of Care	Kessner Index Rating			% of Total Births
	Inadequate, %	Intermediate, %	Adequate, %	
1	1.4	10.6	65.9	77.9
2	1.8	15.7	. . .	17.5
3 or no care	4.6	. . .	. . .	4.6
% of total births	7.7	26.3	65.9	

Note. Percentages are those of all US births. Concordance of trimester and index rating = 86.2%. Percentages may not add to 100 because of rounding.
Source. Percentages are based on an analysis of birth certificate data from the 1980 National Natality Survey.[9]

TABLE 3—Ratings Assigned to Births According to the Number of Visits Component of the Kessner Adequacy of Prenatal Care Index, by the Trimester of Initiation of Prenatal Care

Trimester of Initiation of Care	Number of Visits: Inadequate (<5), %	Intermediate (5–8), %	Adequate (≥9), %	% of Total Births
1	1.4	10.6	65.9	77.9
2	1.8	6.9	8.9	17.5
3 or no care	2.5	1.3	0.8	4.6
% of total births	5.7	18.8	75.5	

Note. Percentages are those of all US births. Concordance of trimester and number of visits = 75.3%. Percentages may not add to 100 because of rounding.
Source. Percentages are based on an analysis of birth certificate data from the 1980 National Natality Survey.[9]

TABLE 4—Ratings Assigned to Births According to the Kessner Adequacy of Prenatal Care Index, Original and Extrapolated beyond Nine Prenatal Care Visits

Rating	Original Kessner	Extrapolated Kessner[a]
Adequate, %	65.9	29.2
Intermediate, %	26.3	54.5
Inadequate, %	7.7	16.3

Note. Percentages are those of all US births; they may not add to 100 because of rounding.
Source. Percentages are based on an analysis of birth certificate data from the 1980 National Natality Survey.[9]
[a]Ratings are based on trimester prenatal care began plus proportion of total visits recommended by American College of Obstetricians and Gynecologists received (all recommended visits received for a rating of Adequate and fewer than half the recommended visits received for a rating of Inadequate).

TABLE 5—Ratings Assigned to Births According to the Kessner Adequacy of Prenatal Care Index, Original and Extrapolated beyond Nine Prenatal Care Visits, by Week of Gestation at Delivery

Week of Gestation at Delivery	Adequate: Kessner, %	Adequate: Extrapolated, %	Intermediate: Kessner, %	Intermediate: Extrapolated, %	Inadequate: Kessner, %	Inadequate: Extrapolated, %
35	57.7	57.7	27.8	27.8	14.5	14.5
36	53.5	53.5	33.1	29.7	13.4	16.8
37	55.7	47.2	31.7	36.6	12.6	16.7
38	61.3	39.8	29.9	44.4	8.8	15.8
39	69.5	38.9	24.6	49.2	5.9	11.9
40	67.0	21.8	26.0	61.5	7.0	16.7
41	73.3	18.5	20.4	68.8	6.4	12.8
42	69.9	13.1	24.6	65.7	5.5	21.1
43	66.8	8.9	23.0	70.4	10.1	20.7
44	66.5	5.5	28.0	68.5	6.6	26.0
45	58.7	3.5	37.4	66.6	3.9	29.9
% of total births	65.9	29.2	26.3	54.5	7.7	16.3

Note. Percentages are those of all US births; they may not add to 100 because of rounding.
Source. Percentages are based on an analysis of birth certificate data from the 1980 National Natality Survey.[9]

required number of visits according to the well-established American College of Obstetricians and Gynecologists (ACOG) recommendations (one visit per month through 28 weeks' gestation, one visit every 2 weeks through 36 weeks' gestation, and one visit per week thereafter).[10] No discussion of the rationale for stopping at nine visits is presented.

It would appear that the index stops at nine visits because only one digit was allocated to record the number of prenatal care visits on the 1968 New York City computerized birth certificate file used in the Kessner et al. analyses.[1] According to ACOG, nine visits corresponds to the recommended number of visits at 36 weeks' gestation; hence the index's adjustments for gestational age stop at that point. Thus, the Kessner Index is constructed on the basis of an outdated computer data storage limitation. Full-term births can, therefore, be rated as having Adequate prenatal care even if they have had fewer than the ACOG-recommended number of visits. The longer the pregnancy, the smaller the percentage of recommended visits needed for care to be rated Adequate. Only premature births (36 weeks' gestation and less) are assessed fully against the ACOG standards.

If one extrapolates the original Kessner Index algorithm beyond the nine visits, using the ACOG standards (e.g., 10 visits by 37 weeks, 11 visits by 38 weeks, etc.), a major redistribution of prenatal care utilization adequacy would occur (Table 4). There would be more than twice the number of women with Inadequate care, a major increase and shift to Intermediate care, and many fewer women with Adequate prenatal care. Indeed, 44% of all women would now be classified as having less adequate prenatal care than previously classified. If one examines the extrapolation stratified by weeks' gestation at delivery, as in Table 5, one sees that this miscategorization increases as gestational age advances.

Finally, the lack of adequate initial documentation for the Kessner Index has led to nonstandardized definitions and discrepancies in its calculations. The algorithm presented in Table 1 appears to be the sole documentation for the Kessner Index. This is clearly inadequate. In particular, there is insufficient description of how to treat records with missing gestational age, missing visits, missing initiation date, etc.[11] The result is that each public health entity has had to program the index itself, with resultant inconsistencies. For example, many states

have added an "unknown" category if number of prenatal care visits or timing of first visit is unknown, while others still follow the original recommendations and record such births as having received "inadequate" care. Some states impute missing gestational age on the basis of birthweight, whereas others (e.g., New York) disregard all such records or disregard only those that cannot be rated as "inadequate" (L. Dellehunt, written communication, September 29, 1986). The state of Missouri and others have further simplified the definition of the Adequate rating of the Kessner Index to eight visits for full-term birth and five visits for premature births (J. W. Stockbauer, written communication, September 26, 1986). Thus, the Kessner Index ratings may not be comparable across sites.

Adequacy of Prenatal Care Utilization Index

The weaknesses of the Kessner Index led the author to try to develop an alternative prenatal care utilization algorithm, based on birth certificate data. The proposed Adequacy of Prenatal Care Utilization (APNCU) Index attempts to characterize prenatal care utilization on two independent and distinctive dimensions: Adequacy of Initiation of Prenatal Care and Adequacy of Received Services (once prenatal care has begun). This two-factor index does not assess the quality of the prenatal care that is delivered, but simply its utilization.

The initial dimension, Adequacy of Initiation of Prenatal Care, characterizes the adequacy of the timing of initiation of care. The month prenatal care begins, which is recorded on the birth certificate, corresponds directly to the adequacy of timing of initiation of prenatal care. The underlying assumption is that the earlier the initiation, the more adequate the prenatal care. The month in which care is initiated is grouped not by trimester, but into four slightly different adequacy groupings: months 1 and 2, months 3 and 4, months 5 and 6, and months 7 through 9. The second trimester was felt to cover too broad a time period to be useful as a single initiation date category. No prenatal care, which can be isolated in this index, is grouped in the late or inadequate care category for this dimension.

The second dimension, Adequacy of Received Services, characterizes the adequacy of the prenatal care visits received during the time the woman is actually in prenatal care (i.e., from initiation until the delivery). The expected number of visits is based on the ACOG prenatal care visitation standards for uncomplicated pregnancies,[12] adjusted for the gestational age at initiation of care and for the gestational age at delivery. The measure for Adequacy of Received Services is the ratio of the actual number of visits to the expected number of visits.

The expected number of visits for each pregnancy can be calculated easily (by computer or by hand) by noting the number of ACOG-recommended prenatal care visits for a pregnancy of a given gestation and then adjusting or reducing that number based on the gestational age at initiation of care (missed visits are assumed not to be made up). Essentially, this procedure measures a segment of a fixed recommended prenatal care utilization metric, but in contrast to the Kessner Index, it is adjusted at two places: when the woman begins prenatal care and when she delivers. For example, for a 40-week pregnancy, ACOG recommends 14 visits; if care began in month 4 (three missed visits), then the expected number of visits would be 11. The number of actual or observed visits can be directly recorded from the birth certificate (or any other prenatal care data source). The ratio of observed to expected visits is then grouped into four categories: Inadequate (less than 50% of expected visits), Intermediate (50%–79%), Adequate (80%–109%), and Adequate Plus (≥110%). A similar ratio concept is implicit in the existing Kessner Index, wherein inadequate visits equal approximately 50% of adequate visits. These four percentage categories allow for a slightly broader range of numbers of visits to be rated as Adequate care (80%–109%) and provide, for the first time, a measurement of prenatal care utilization that exceeds ACOG standards. This dimension of Adequacy of Received Services is independent of the previously described dimension of Adequacy of Initiation of Prenatal Care.

The two dimensions can be combined into a single summary prenatal care utilization index. Inadequate utilization is defined as either late initiation (after the 4th month of pregnancy) or less than 50% of recommended visits. All other categories require initiation of care by the 4th month of pregnancy and then are coded according to the extent of received services (e.g., to be rated Adequate Plus requires initiation of care between 1 and 4 months and more than 110% of the expected ACOG-recommended visits, etc.). Figure 1 portrays the construction of the summary APNCU Index.

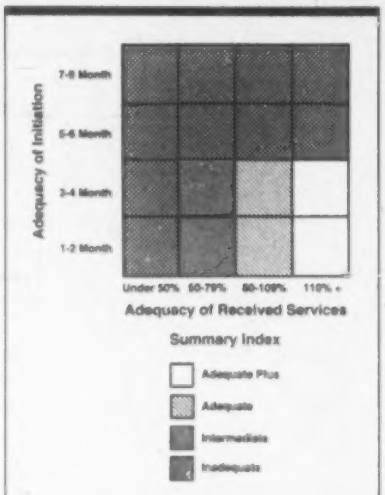


FIGURE 1—The summary Adequacy of Prenatal Care Utilization Index.

A descriptive outline of the proposed index and its two factors is presented in Table 6. (A more detailed description of the APNCU Index and its coding assumptions and an SAS Program for its computation are available from the author.)

Table 7 presents a comparison of the APNCU and Kessner Index ratings. Only 71.5% of women in the United States would be rated the same on the two indexes (assuming the APNCU's "Adequate and Adequate Plus" equal Kessner's "Adequate"), with 21.1% achieving a poorer rating and 7.4% an improved rating. Women whose care was rated Intermediate on the Kessner Index would be the most likely to be recategorized: 34% would now be rated Inadequate and 28% rated Adequate or Adequate Plus.

In a related paper in this issue of the Journal I apply the APNCU Index to the 1980 National Natality Survey data to assess the adequacy of prenatal care utilization in the United States and its association with low birthweight.[13]

Discussion

The assessment of the adequacy of prenatal care utilization is heavily shaped by the way in which utilization is measured. One of the major strategic and programmatic thrusts to reduce infant mortality in this era is to increase early initiation of prenatal care. *Healthy People*

TABLE 6—Outline of the Adequacy of Prenatal Care Utilization Index

I. Month prenatal care began (Adequacy of Initiation of Prenatal Care)
   - Adequate Plus: 1st or 2nd month
   - Adequate: 3rd or 4th month
   - Intermediate: 5th or 6th month
   - Inadequate: 7th month or later, or no prenatal care

II. Proportion of the number of visits recommended by the American College of Obstetricians and Gynecologists received from the time prenatal care began until delivery (Adequacy of Received Services)
   - Adequate Plus: ≥ 110%
   - Adequate: 80–109%
   - Intermediate: 50–79%
   - Inadequate: < 50%

III. Summary Adequacy of Prenatal Care Utilization Index
   - Adequate Plus: Prenatal care begun by the 4th month and 110% or more of recommended visits received
   - Adequate: Prenatal care begun by the 4th month and 80%–109% of recommended visits received
   - Intermediate: Prenatal care begun by the 4th month and 50%–79% of recommended visits received
   - Inadequate: Prenatal care begun after the 4th month or less than 50% of recommended visits received

TABLE 7—Ratings Assigned to Births According to the Adequacy of Prenatal Care Utilization Index Compared with the Kessner Index

Kessner Index	Adequacy of Prenatal Care Utilization Index: Inadequate, %	Intermediate, %	Adequate, %	Adequate Plus, %	% of Total Births
Inadequate	7.7	0.0	0.0	0.0	7.7
Intermediate	8.9	10.1	5.0	2.4	26.3
Adequate	.1	12.1	38.3	15.4	65.9
% of total births	16.7	22.2	43.4	17.7	100.0

Note. Percentages are those of all US births; they may not add to 100 because of rounding.
Source. Percentages are based on an analysis of birth certificate data from the 1980 National Natality Survey.[8]

2000's goal 14.11 is to increase first-trimester prenatal care to at least 90% of live births.[14] By contrast, continuity of prenatal care once enrolled is much less emphasized. Professional and public attention has been drawn to the theme of early access to care, I believe, because we have been able to marshal clear evidence about prenatal care initiation but we lack any readily available measures of care after enrollment. Both current popular measures of prenatal care adequacy, trimester of initiation of care (by definition) and the Kessner Index (by algorithmic biases), are basically measures of initial access to care. Improving birth outcomes, however, may be dependent on other features of prenatal care (such as content, timing, and number of visits). One utilization measure may not capture all facets of prenatal care.

The Kessner Index was a major achievement in perinatal health service research, transforming two technically available but continuous and complex data items on birth certificates into a simple three-point utilization scale. Its original rationale and basic algorithm seemed clear and clinically reasonable. Unfortunately, the Kessner Index appears to be seriously flawed; it may be leading us to misperceive the nature of prenatal care utilization in the United States. Four limitations were noted in this paper.

First, the Kessner Index is overwhelmingly a measure of the initiation of prenatal care. Only 14% of women receive fewer visits than the number required for the trimester they enter care. This initiation bias may explain why those who have used both the Kessner Index and the trimester of prenatal care initiation note so little difference between them in most analyses; the latter measure is therefore preferred because it does not require the often inaccurately recorded or missing gestational age variable in its calculations (J.C. Kleinman, PhD, verbal communication, November 6, 1987). Second, the Kessner Index does not distinguish between inadequacy due to late initiation and inadequacy due to an inadequate number of visits. Although the summary Kessner Index was not developed to measure these components separately, the absence of independent measures results in the loss of important information about the nature of prenatal care adequacy, especially since 24% of women would be rated differently on these two dimensions. Neither of these first two observations about the Kessner Index, through interesting, would seem to warrant its dismissal.

The limitation of the Kessner Index to nine visits is, however, critical. This limitation is totally arbitrary and not clinically derived; it is the direct result of a computer data capacity limitation of the 1968 New York City birth file. For 20 years, the US public health profession has based its major index of prenatal care adequacy on an algorithm developed to accommodate this single-digit limitation in the counting of the number of prenatal care visits.

Because of the nine-visit limitation, the Kessner Index incorrectly assesses prenatal care utilization adequacy for normal and post-term births, the vast majority of births in the United States. The extrapolated Kessner Index algorithm would indicate that only 29%, not 66%, of births receive "adequate" care. This is not a minor difference in our perception of the extent of prenatal care adequacy in the United States. Accurate assessment of prenatal care utilization for term infants may be particularly important, given the significant racial disparities in birth outcomes for normal-birthweight infants[15] and the recent observations that there are significant Black–White differences in the utilization of prenatal care at the end of pregnancy.[16,17] Moreover, the limitation to nine visits also biases the assessment of the relationship of prenatal care and birth outcomes. Full-term babies, which are bigger and more frequently healthy, are more readily rated as having received "adequate" care than are preterm babies, thus artificially enhancing the association of positive birth outcomes with more positive prenatal care adequacy ratings.[18] This bias suggests that the

current literature using the Kessner Index may be overstating the limited but positive association previously noted.[3,19]

Finally, the lack of documentation for the Kessner Index, not surprising in a less computer-intensive era, has resulted in different computations of the Kessner Index in different localities. Alexander et al.[11] have shown in great detail how different coding conventions regarding missing data can result in major differences in the measurement of prenatal care adequacy and the evaluation of perinatal care programs such as Medicaid.

The proposed APNCU Index attempts to correct these four limitations of the Kessner Index. The independent assessment of prenatal care utilization after initiation, adjusted for the full range of gestational ages, is clearly the most important new feature of the APNCU Index. It provides information and a perspective on prenatal care utilization that is not presently available to the US public health community. This is important because many maternal and child health intervention programs are targeted toward continuity of prenatal care services or enhancement of services once a woman has entered care (e.g., case management, risk screening, home visitations). This new component should allow for a more direct assessment of these initiatives, independent of the timing of initiation of care. Recent Medicaid enhancements appear to have differential effects on the two different components of prenatal care utilization.[20,21]

The establishment of an Adequate Plus category, another innovative feature of the APNCU Index, provides a means to directly estimate the number of women receiving more than the ACOG-recommended number of visits, adjusted for the timing of care initiation. It appears that it is important to isolate this fairly large group of high-risk women because they have a disproportionate number of the low-birthweight babies.[13] Efforts to isolate this high-risk group based simply on an absolute number of visits (13+) will not be accurate, incorrectly classifying many women who start prenatal care late or who are post-term.

Though not emphasized in the data presentations in this paper, the APNCU Index does allow for the direct ascertainment of the extent of "no prenatal care" as a subcategory of Inadequate prenatal care. Several perinatal analysts have emphasized the importance of distinguishing between these two groups.[22,23]

The proposed APNCU Index can be seen as a second-generation prenatal care adequacy index. The APNCU Index was developed in 1987 and improved through feedback from multiple early users. The current version (III) includes an Adequate Plus category in the Summary Index; distinguishes "no prenatal care" within the Inadequate care category; has further clarified the coding of "no data," "missing data," and unusual data combinations; and allows users to adapt the computer program for unusual coding conventions in their own databases. The present version is basically a very minor revision of the prior APNCU Index's algorithms.

There are limitations to this new APNCU Index. First, it does not measure the adequacy of the content of prenatal care, but merely that of the utilization of prenatal care; hence its name. Second, the APNCU Index is only as accurate as the data (birth certificate or otherwise) used to calculate it. Inaccuracies in birth certificate data have been well documented, especially for prenatal care information[2] and gestational age.[24] Third, this index has the opposite bias of the Kessner Index: the longer the pregnancy, the more opportunity to miss prenatal care visits, and hence the less likelihood of a rating of Adequate or Adequate Plus.[18] This bias is not, however, built into the structure of the index; the adequacy ratings accurately reflect the increasing difficulty that women have in meeting the demanding ACOG recommendations as the pregnancy continues. Fourth, the present APNCU Index does not adjust for the risk conditions of the mother. The ACOG recommendations are for women with uncomplicated pregnancies; thus, the APNCU Index produces a slightly conservative estimate of inadequate prenatal care utilization because it underestimates the true need for prenatal care visits.

Conclusions

The proposed APNCU Index, with its two components, provides a more accurate and comprehensive set of measures of prenatal care utilization than the widely used Kessner Index. The accurate measurement of prenatal care utilization is a critical step in the development of public health programs to improve prenatal care services and ultimately to improve birth outcomes. □

Acknowledgments

An earlier version of this paper was presented at the 115th Annual Meeting of the American Public Health Association, New Orleans, La, October 18–21, 1987.

The author wishes to thank Paul Wise, Chris Costello, Diane Ramsey, and the many users of early versions of the APNCU Index who made suggestions to improve its versatility and accuracy.

References

1. Kessner DM, Singer J, Kalk CE, Schlesinger ER. *Infant Death: An Analysis by Maternal Risk and Health Care.* Washington, DC: Institute of Medicine and National Academy of Sciences; 1973:chap 2.
2. Forrest JD, Singh S. Timing of prenatal care in the United States: how accurate are our measurements? *Health Serv Res.* 1987; 22:235–253.
3. Liberatos P, Kiely JL. Selected issues in the evaluation of prenatal care. In: Kiely M, ed. *Reproductive and Perinatal Epidemiology.* Boca Raton, Fla: CRC Press; 1991: 79–98.
4. Gortmaker SL. The effects of prenatal care upon the health of the newborn. *Am J Public Health.* 1979;69:653–660.
5. Quick JD, Greenlick MR, Roghmann KJ. Prenatal care and pregnancy outcome in an HMO and general population: a multivariate cohort analysis. *Am J Public Health.* 1981;71:381–390.
6. Showstack JA, Budetti PP, Minkler D. Factors associated with birthweight: an exploration of the roles of prenatal care and length of gestation. *Am J Public Health.* 1984;74:1003–1008.
7. Murray JL, Bernfield M. The differential effect of prenatal care on the incidence of low birthweight among whites and blacks in a prepaid health care plan. *N Engl J Med.* 1988;319:1385–1391.
8. Placek PJ. The 1980 National Natality and Fetal Mortality Surveys. *Public Health Rep.* 1984;99:111–116.
9. Keppel KG, Heuser RL, Placek PJ. Methods and response characteristics, 1980 National Natality and Fetal Mortality Surveys. *Vital Health Stat* [2]. 1986; no. 100. DHHS publication PHS 86-1374.
10. *Manual of Standards in Obstetric-Gynecologic Practice.* 2nd ed. Chicago, Ill: American College of Obstetricians and Gynecologists; 1965.
11. Alexander GR, Tompkins ME, Petersen DJ, Weiss J. Sources of bias in prenatal care utilization indices: implications for evaluating the Medicaid expansion. *Am J Public Health.* 1991;81:1013–1016.
12. *Standards for Obstetric-Gynecologic Services.* 6th ed. Washington, DC: American College of Obstetricians and Gynecologists; 1985.
13. Kotelchuck M. The adequacy of prenatal care utilization index: its US distribution and association with low birthweight. *Am J Public Health.* 1994;84:1486–1489.
14. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91-50212.
15. Hogue CJ, Buehler JW, Strauss LT, Smith

JC. Overview of the national infant mortality surveillance project—design, method and results. *Public Health Rep.* 1987;102: 126–138.
16. Kotelchuck M, Costello CA, Wise PH. Race differences in utilization of prenatal care services in the U.S. Presented at the 115th Annual Meeting of the American Public Health Association; October 18–21, 1987; New Orleans, La.
17. Kogan MD, Kotelchuck M, Johnson S. Racial differences in late prenatal care visits. *J Perinatol.* 1993;13:14–21.
18. Kleinman JC. Methodological issues in the analysis of vital statistics. In: Kiely M, ed. *Reproductive and Perinatal Epidemiology.* Boca Raton, Fla: CRC Press; 1991:447–468.
19. Institute of Medicine. *Preventing Low Birthweight.* Washington, DC: National Academy Press; 1985.
20. Braveman P, Bennett T, Lewis C, et al. Access to prenatal care following major Medicaid eligibility expansions. *JAMA.* 1993;269:1285–1289.
21. Krieger JW, Connell FA, LoGerfo JP. Medicaid prenatal care: a comparison of use and outcomes in fee-for-service and managed care. *Am J Public Health.* 1992;82: 185–190.
22. Harris JE. Prenatal medical care and infant mortality. In: Fuchs VR, ed. *Economic Aspects of Health.* Chicago, Ill: University of Chicago Press; 1982:15–52.
23. Alexander GR, Cornely DA. Prenatal care utilization: its measurement and relationship to pregnancy outcome. *Am J Prev Med.* 1987;3:243–253.
24. Fingerhut A, Kleinman JC. Comparability of reporting between the birth certificate and the 1980 National Natality Survey. *Vital Health Stat* [2]. 1985; no. 99. DHHS publication PHS 86-1373.

Call for Papers for New Journal of Sudden Infant Death Syndrome and Infant Mortality

Although sudden infant death syndrome (SIDS) is the major cause of infant death after the first 2 weeks of age, there is no journal devoted to SIDS. The *Journal of Sudden Infant Death Syndrome and Infant Mortality* will fill this gap by providing an international forum for the presentation and discussion of information that contributes to the understanding of SIDS and of other unexpected deaths in the vulnerable postneonatal period. This quarterly, peer-reviewed publication will include original articles on basic research and clinical studies, scholarly reviews, symposia proceedings, and book reviews. To support the goal of understanding the pathophysiology of SIDS, the journal will be concerned with new information about normal infant development, as well as with broad aspects of unexpected infant deaths. In addition, the journal will present articles that explore bereavement and counseling issues.

The *Journal of Sudden Infant Death Syndrome and Infant Mortality* is intended for researchers, clinicians, other health care professionals, and families affected by SIDS who will benefit from up-to-date information on the progress of SIDS research and prevention. The journal's editor-in-chief is Lois M. Roeder, director of the Professional Training and Education Unit, Sudden Infant Death Syndrome Institute, Department of Pediatrics, University of Maryland School of Medicine. The journal will be published under the auspices of the Sudden Infant Death Syndrome Institute at the University of Maryland, Baltimore.

For submission information, contact Lois M. Roeder, ScD, Editor-in-Chief, SIDS Journal, SIDS Institute, 22 S Greene St, Room N5W67, Baltimore, MD 21201-1595.

For subscription information, contact Journals Customer Service, Plenum Publishing Corporation, 233 Spring St, New York, NY 10013; tel (800) 221-9369 or (212) 620-8470/71.

Prenatal Alcohol Exposure, Birthweight, and Measures of Child Size from Birth to Age 14 Years

Paul D. Sampson, PhD, Fred L. Bookstein, PhD, Helen M. Barr, MA, MS, and Ann P. Streissguth, PhD

ABSTRACT

Objectives. The purpose of the study was to examine the effect of prenatal alcohol exposure on offspring's weight, height, and head circumference from birth through 14 years of age.

Methods. This longitudinal prospective study examined a cohort of approximately 500 offspring (oversampled for heavier drinkers and stratified for smoking from a population of 1529 women in prenatal care at the 5th gestational month) at birth; 8 and 18 months; and 4, 7, and 14 years of age. Covariates were examined by means of multiple regression. Birth size measures were also examined as predictors of 7-year neurodevelopmental outcomes.

Results. Effects of alcohol were observed on weight, length, and head circumference at birth; these effects were not altered by adjustment for covariates including smoking. However, the birthweight effect is clearly transient: although alcohol effects remained observable at 8 months, they were not measurable thereafter through age 14 years.

Conclusions. In this population-based sample, neither birthweight nor any later size measure was as useful an indicator of the enduring effects of prenatal alcohol exposure as were certain neurodevelopmental outcomes. (*Am J Public Health.* 1994; 84:1421–1428)

Introduction

There is a substantial literature on the association between maternal alcohol consumption during pregnancy and decreased neonatal weight, length, and head circumference. Indeed, one of the three defining characteristics of fetal alcohol syndrome is prenatal onset growth deficiency. However, conflicting results have been reported across studies of alcohol consumption at less than extreme levels (levels not as high as frank maternal alcoholism). Among the prospective studies that have measured neonatal size directly, several have found effects of alcohol exposure on birthweight (effects varying in magnitude and, in at least one case, depending on whether or not the mother smoked).[1–14] Other studies have not.[15–20] As Zuckerman and Hingson noted, these studies vary widely in sample ranges and variances of dose, in the timing and patterns of exposure, in the nature of the dose measurement(s) chosen, in the protocol for that assessment, and in the distributions of, and correction for, other factors affecting neonatal size.[21]

Regarding long-term consequences of fetal alcohol exposure for child size, Day et al. found increasing effects through age 36 months[22–24]; Russell et al. found height and head circumference effects at 6 years[25]; and Coles et al. found effects only on head circumference at 5½ years.[26] On the other hand, O'Connor et al. found no size effects at 12 months[27]; Fried and O'Connell found no effects at 12 or 24 months[10]; and Greene et al. found no conclusive evidence for persistence of effects through the preschool years (through age 4 years, 10 months).[13]

In this paper we attempt to clarify this literature from the perspective of our analyses of data from the Seattle Longitudinal Prospective Study on Alcohol and Pregnancy. These analyses permit us to follow the effects of prenatal exposure (at moderate levels) on children's growth in a sample that is reasonably balanced with respect to the principal competing teratogenic hypothesis—smoking-related deficiency of birthweight—and that offers multiple measures of dose. There is no doubt that alcohol doses in this range have substantial, enduring effects on the neurobehavioral development of the child through 14 years of age.[27–34] In this paper we ask whether these neurobehavioral effects are related to the effects of alcohol on size. Behavioral teratology tenets would suggest they need not be.[35]

We find that there is a distinct effect of prenatal alcohol exposure on measures of size at birth that can be detected with variable sensitivity by a variety of dose measures. The effect is shared across height, weight, and head circumference. After the age of 8 months the effects on size are essentially lost. Furthermore, the amount of apparent neonatal growth deficiency does not correlate with or

Paul D. Sampson is with the Department of Statistics, Helen M. Barr is with the Department of Psychiatry and Behavioral Sciences, and Ann P. Streissguth is with the Department of Psychiatry and Behavioral Sciences, the Child Development/Mental Retardation Center, and the Alcoholism and Drug Abuse Institute, all at the University of Washington, Seattle. Fred L. Bookstein is with the Center for Human Growth and Development, University of Michigan, Ann Arbor.

Requests for reprints should be sent to Ann P. Streissguth, PhD, Department of Psychiatry and Behavioral Sciences, GG-20, University of Washington, Seattle, WA 98195.

This paper was accepted February 23, 1994.

Editor's Note. See related editorial by Paneth (p 1372) in this issue.

TABLE 1—Descriptive Statistics for Prenatal Alcohol Scores for the 436 of 553 Mothers Who Did Not Abstain Just Prior to and during Pregnancy

	Mean	SD	Minimum	Median	Maximum
Average ounces of absolute alcohol per day					
Prior to pregnancy recognition	0.74	1.61	0	0.35	25.76[a]
During pregnancy	0.33	0.63	0	0.15	8.55
Quantity-frequency-variability index					
Prior to pregnancy recognition	3.35	1.23	1	3	5
During pregnancy	2.99	0.99	1	3	5
Binge drinking (5 drinks or more on any occasion)					
Prior to pregnancy recognition	0.36	0.48	0	0	1
During pregnancy	0.24	0.43	0	0	1
Maximum no. of drinks reported on any occasion					
Prior to pregnancy recognition	3.77	2.79	0	3.50	13.00
During pregnancy	3.49	2.61	0	3.50	13.00
Monthly occasions of drinking					
Prior to pregnancy recognition	15.62	21.67	0	1.70	5.48
During pregnancy	8.40	12.85	0	0.92	4.80
Average no. of drinks per occasion					
Prior to pregnancy recognition	2.32	1.57	0	0.92	2.64
During pregnancy	2.12	1.18	0	0.92	2.64

Note. All skewed distributions were log-transformed for analysis.
[a]This score was an extreme outlier; the next highest score was 9.14 ounces.

predict later neurobehavioral consequences of the prenatal exposure.

Methods

The Seattle Longitudinal Prospective Study on Alcohol and Pregnancy began in 1974 when a consecutive sample of pregnant women planning to deliver at either of two hospitals (n = 1529) were interviewed during their 5th month of pregnancy regarding their smoking, drinking, and drug use during pregnancy. A follow-up cohort of approximately 500 children was selected at birth, oversampled for heavier alcohol exposure and stratified for smoking and alcohol in order to include children born to mothers ranging from heavy drinkers to infrequent drinkers and abstainers and to include both smokers and nonsmokers at all levels. (For further details see Streissguth et al.[29,33,36]) Both the prenatal sample and the follow-up cohort were primarily White, married, and middle class. Approximately 80% of the women reported drinking during pregnancy, but fewer than 1% reported any serious problems with drinking.[37] Maintenance of this follow-up cohort over time has been excellent. Only 18% were lost to follow-up by 14 years (primarily because they moved out of the area). At the 7-year examination we saw 95% of the children examined at 4 years, and at the 14-year examination we saw 93% of the children examined at 7 years. Complete weight data for all six examination periods is available on 335 children. The maximum number of children examined in the present analyses is 553, representing all children in the cohort who were seen in at least one of the follow-up examinations. Delivery data on weight, height, and head circumference were abstracted from the medical records of 1439 of the original 1529 mothers interviewed during midpregnancy.[38] Subsequent measurements were obtained during laboratory visits.

The prenatal interview included questions about all medications and other drugs used during pregnancy, smoking and dietary habits, and other life-style and pregnancy history factors thought to relate to pregnancy outcome.[33,36] The quantity-frequency-variability structured interview[39] was given for beverage consumption during two time periods: prior to pregnancy or recognition of pregnancy and during pregnancy. Separate questions were asked for beer, wine, liquor, and cola beverages. The questions about alcoholic beverages yielded a variety of derived scores reflecting variations in level, timing, and pattern of prenatal alcohol exposure. The quantitative scores—average ounces of alcohol per day, number of drinking occasions monthly, average number of drinks per drinking occasion, and maximum drinks on any occasion—all had highly skewed distributions and were log-transformed for analysis. We also included a 5-point scaled alcohol score of quantity-frequency-variability proposed by Cahalan et al.,[39] a binary indicator of drinking five or more drinks on at least one occasion, and an indicator of complete abstinence from alcohol (prerecognition and during pregnancy). Distributions of scores on our 12 primary measures of alcohol exposure (prior to log transformation) are summarized in Table 1 for drinkers only. The precision of these self-report measures is discussed in a previous publication.[37] The correlations among the alcohol measures based on all 553 cases range from .39 to .91. Average daily nicotine use was calculated by multiplying the number of cigarettes smoked per day by the nicotine content of the brand smoked. (The logarithms of the number of cigarettes smoked [+1] and nicotine [+1] are highly correlated in this sample; $r = .85$.)

Analyses and Results

In our sample, in spite of its having been overweighted for heavier drinkers, low-birthweight babies (those weighing less than 2500 g) per se were rare. Our screening sample included 41 infants out of 1437 survivors with birthweights below 2500 g (2.9%). In our follow-up sample, only 12 of 553 infants (2.1%) were below this threshold. By contrast, National Center for Health Statistics standards suggest a low-birthweight incidence of 5% in neonates of normal, healthy, middle-class White mothers,[40] and Day et al. found an incidence of 11% in their low-socioeconomic-status sample.[12]

Our statistical analyses are presented in three stages: (1) simple correlations of alcohol measures with weight, height, and head circumference at six ages from birth to 14 years; (2) covariate adjustment with regression analysis; and (3) individual growth trajectories of specific children.

Simple Correlations

Table 2 presents the correlations of each of our measures of prenatal alcohol dose with the three measures of child size at six ages. Correlations for size measures at delivery are based on slightly larger sample sizes than are the follow-up measures, but the results are little affected by the subjects lost to follow-up.

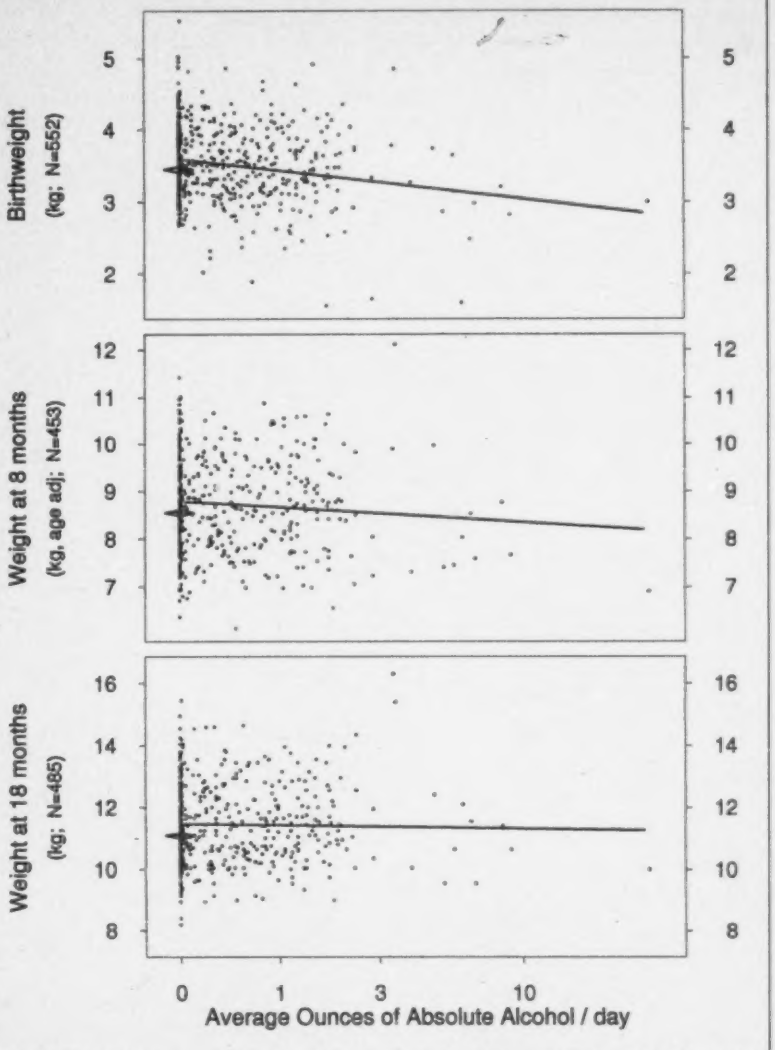
TABLE 2—Correlations of Measures of Prenatal Alcohol Exposure with Growth through 14 Years of Age

	Delivery	8 mo	18 mo	4 y	7 y	14 y
Weight	(n = 552)	(n = 453)	(n = 485)	(n = 451)	(n = 481)	(n = 452)
Average ounces of absolute alcohol per day[a]						
Prior to pregnancy recognition	−.15	−.05	.03	.01	.01	.05
During pregnancy	−.09	−.06	−.02	.01	−.01	−.02
Quantity-frequency-variability index						
Prior to pregnancy recognition	−.12	−.03	.06	.03	.04	.02
During pregnancy	−.06	−.01	.07	.02	.00	.00
Binge drinking (5 drinks or more on any occasion)						
Prior to pregnancy recognition	−.15	.01	.05	−.03	.02	−.01
During pregnancy	−.11	.01	.05	.02	.04	.06
Maximum no. of drinks reported on any occasion[a]						
Prior to pregnancy recognition	−.11	−.00	.08	.03	.03	.02
During pregnancy	−.04	.02	.07	.01	.03	.01
Monthly occasions of drinking[a]						
Prior to pregnancy recognition	−.09	−.02	.04	.03	.02	.01
During pregnancy	−.04	−.01	.03	.04	−.01	−.03
Average no. of drinks per occasion[a]						
Prior to pregnancy recognition	−.11	−.03	.06	.04	.04	.03
During pregnancy	−.00	.04	.09	.03	.04	.02
Length/height	(n = 470)	(n = 453)	(n = 479)	(n = 451)	(n = 482)	(n = 457)
Average ounces of absolute alcohol per day[a]						
Prior to pregnancy recognition	−.17	−.03	.02	.05	.02	.07
During pregnancy	−.08	−.05	−.02	.06	.08	.07
Quantity-frequency-variability index						
Prior to pregnancy recognition	−.08	.02	.09	.07	.08	.08
During pregnancy	−.06	.00	.05	.06	.04	.07
Binge drinking (5 drinks or more on any occasion)						
Prior to pregnancy recognition	−.10	.05	.08	.03	.04	.04
During pregnancy	−.06	.02	.04	.02	−.01	.05
Maximum no. of drinks reported on any occasion[a]						
Prior to pregnancy recognition	−.07	.05	.10	.07	.07	.07
During pregnancy	−.03	.03	.04	.02	.04	.03
Monthly occasions of drinking[a]						
Prior to pregnancy recognition	−.10	.01	.07	.08	.07	.09
During pregnancy	−.05	−.01	.03	.08	.09	.09
Average no. of drinks per occasion[a]						
Prior to pregnancy recognition	−.06	.03	.09	.08	.07	.07
During pregnancy	−.00	.06	.06	.05	.04	.04
Head circumference	(n = 536)	(n = 453)	(n = 485)	(n = 451)	(n = 475)	(n = 455)
Average ounces of absolute alcohol per day[a]						
Prior to pregnancy recognition	−.11	−.03	.02	.02	.02	.00
During pregnancy	−.06	−.06	−.02	−.00	.03	−.03
Quantity-frequency-variability index						
Prior to pregnancy recognition	−.05	−.04	.01	.02	.02	−.01
During pregnancy	−.01	−.03	.03	.01	.01	.00
Binge drinking (5 drinks or more on any occasion)						
Prior to pregnancy recognition	−.07	.01	−.01	−.02	−.01	−.03
During pregnancy	−.08	−.02	−.02	−.02	−.04	−.04
Maximum no. of drinks reported on any occasion[a]						
Prior to pregnancy recognition	−.04	−.03	.01	.01	.00	−.01
During pregnancy	.02	−.05	.01	−.00	−.02	−.01
Monthly occasions of drinking[a]						
Prior to pregnancy recognition	−.05	−.00	.05	.07	.06	.03
During pregnancy	−.02	−.01	.05	.06	.07	.03
Average no. of drinks per occasion[a]						
Prior to pregnancy recognition	−.03	−.05	−.00	−.00	−.01	−.01
During pregnancy	.06	−.03	.03	.02	.00	.02

Note. All correlations are unadjusted except at 8 months. A sampling bias, inadvertently introduced into the 8-month data by examining babies of heavier-drinking mothers at a significantly older age, necessitated the use of age-adjusted correlations provided in this column.
[a]Log-transformed scores were used for analysis because of skewed distributions.

The 8-month data have been linearly adjusted for age of the child at the 8-month examination because the children who were examined relatively late (at this examination only) were disproportionately the offspring of heavier drinkers.

At birth, there is a consistent pattern of negative correlation for all dose measures against all three size measures. At 8



Note. The horizontal open diamond superimposed on the vertical row of abstainers in each panel indicates the level of the abstainer effect.

FIGURE 1—Scatterplots of average ounces of absolute alcohol per day prior to pregnancy recognition (AAP), by birthweight (top), age-adjusted weight at the 8-month examination (center), and 18-month weight (bottom).

months this pattern is still apparent, but the magnitude of the correlations is not as strong. After 8 months, the consistency of these signs wholly vanishes. The neonatal size measure having the highest average correlation with this suite of dose measures is birthweight; of the 12 alcohol scores, the one having the highest average correlation with the suite of three size measures is log average ounces of absolute alcohol per day prior to pregnancy recognition (AAP). After birth, only head circumference seems to display any pattern of consistent (although modest) correlation with alcohol, but in this case only for binge-oriented measures such as average number of drinks per drinking occasion and maximum number of drinks per occasion, and only after deletion of or adjustment for abstainers. The resulting correlations with these binge scores remain in the range −.10 to −.12 over the last three examination periods.

Figure 1 presents scatterplots of AAP by birthweight, weight at 8 months, and weight at 18 months. These scatters represent correlations of −.15, −.05, and −.03, respectively, for the three ages as reported in Table 2. The figure displays fitted regression lines computed only for the mothers who drank. Horizontal open diamonds mark the mean weights of the children of abstainers. Correlations for the nonabstainers are −.20, −.09, and −.02, respectively. The effect of alcohol on birthweight is evident whether or not abstainers are included. At 8 months, the smaller alcohol effect is detectable only when abstainers are excluded. At 18 months and beyond, no alcohol effects are evident.

The association suggested by the figure can be summarized by dividing the sample according to four dose ranges calibrated by AAP: the 154 children of abstainers (AAP of 0); the 125 with positive AAP of less than 0.14; the 18 children whose doses were greater than 4.5 drinks per day on average (AAP greater than 2.28; the top 1/32 of the sample); and the remaining 261 children. At birth, the most substantial effect is a contrast of the highest-exposure group with the low-exposure group (excluding the abstainers). The mean weights in grams and standard deviations (SDs) for these four groups, from highest-exposure to abstainers, are as follows: 3184 (795), 3465 (516), 3561 (474), and 3518 (514). At 8 months the most striking dependence of weight upon dose appears to be a nonlinear tail effect, with a decrease in mean weight evident only in the highest-exposure group. The mean weights in grams and standard deviations, from highest-exposure group to abstainers, are now 8386 (1370), 8715 (961), 8632 (985), and 8696 (973). In fact, the offspring of abstainers are slightly lighter in mean weight than the most highly exposed cases at all later ages. We suspect that this effect is not teratological but sociological.

A multivariate analysis of all exposure and outcome variables together was carried out, using nonlinear partial least squares methods described in Sampson et al.[30] This latent variable analysis does not reveal any long-term relationship between overall alcohol exposure, represented as a composite linearized dose measure, and child size taken across the lifespan from birth to 14 years. The multivariate analysis thus supports the data from the table of simple correlations (Table 2). Prenatal alcohol shows its strongest relationship with birthweight; is still associated with 8-month weight, although the magnitude of the effect is smaller; and rapidly falls to the noise level by 18 months. This phenom-

enon (the vanishing of a correlation with size over time) is a likely consequence of what is normally referred to as "catch-up growth." The correlations of dose with changes in weight are in accordance with what one expects from the correlations wave by wave and supply no additional insights.

Covariate Adjustment

A wide variety of factors correlate with birthweight (cf. Goldstein[41]). In Table 3 we display correlations of a variety of well-known determinants of child size against the six outcome measures for weight. (Tables of correlations with the height and head circumference measures are similar.) Most predictors fall into two groups: those having effects that remain substantial over time (e.g., sex and maternal size measures) and those having correlations with size that quickly fall to zero after birth (e.g., alcohol, nicotine, marijuana, and prematurity).

In extensive multiple regression analyses not reported here, all the variables presented in Table 3 were considered in assessments of the significance of prenatal alcohol exposure as a predictor of child size. Most models used log AAP as the primary predictor of size because it is the alcohol measure of highest average correlation with all the birth size measures. In these regressions, child size was primarily predicted by maternal size (height, pregnancy weight, and weight gain during pregnancy) and sex of the child. None of the nicotine, marijuana, and alcohol scores were reliable predictors of child size after 8 months. These conclusions were unchanged when the least significant terms were dropped from the regression models. In addition, these conclusions were unaffected by optimally transforming the alcohol scores (as in Sampson et al.[30]) or by replacing them with a composite, such as a latent alcohol variable reflecting the joint impact of all the measured dimensions of alcohol dose represented in the 12 prenatal alcohol scores.

Regression analyses using the binge alcohol measures, which correlate best with postnatal head circumference, show similar results: consistently modest, but not significant, estimated effects on head circumference, as suggested by the simple correlations in Table 2.

Individual Growth Trajectories

Separately for each size measure, we considered the babies who were among the lowest 15% on that measure and among the top 15% on AAP. There were 19, 15, and 11 babies who met the criteria for weight, length, and head circumference, respectively. Even these low-birth-size/heavier-exposure babies, in general, showed increased catch-up growth on height, weight, and head circumference, relative to the rest of the cohort, after 8 or 18 months. Infants selected only for extreme size demonstrated, on average, a similar degree of catch-up growth, or what might be called "regression to the mean."

We also examined growth trajectories across time for the three children in this study who were diagnosed with fetal alcohol syndrome by dysmorphologists (two at birth[42] and one, who was not examined at birth, at 4 years[43]). As growth deficiency is one of the characteristics of fetal alcohol syndrome, one would not expect catch-up growth in persons with this diagnosis, at least until the onset of puberty. But neither of the two children diagnosed with the syndrome at birth was excessively low in birthweight. One was more than 2 standard deviations below our sample mean for length, and both were near −2 standard deviations for head circumference. Even these children with a clinical diagnosis of fetal alcohol syndrome showed increasingly normal weight and height over time. Head circumference was more consistently low. The three children with clinical fetal alcohol syndrome showed a continued postnatal decrement in head circumference, relative to the rest of the sample, through 8 to 18 months of age.

TABLE 3—Correlations of Prenatal Demographic Characteristics and Covariates with Weight through 14 Years of Age

	Delivery	8 mo	18 mo	4 y	7 y	14 y
Preterm gestational age	.45	.13	.06	−.02	−.04	−.06
Postterm gestational age	.30	.12	.08	.00	−.02	−.05
Maternal age	.11	−.07	−.01	.01	−.03	−.05
Maternal height	.16	.25	.32	.32	.28	.14
Maternal prepregnancy weight	.22	.26	.34	.36	.39	.35
Maternal delivery weight	.31	.30	.36	.37	.43	.35
Maternal weight gain	.22	.13	.07	.06	.11	.02
Prenatal nutrition	.07	.01	−.01	−.02	−.08	−.14
Mother abstained during pregnancy	−.04	−.06	−.12	−.06	−.05	−.02
Log parity	.10	−.15	−.07	−.08	−.06	−.05
Mother used aspirin	.03	.10	.12	.07	.06	.10
Mother used acetaminophen	.01	−.01	.02	.03	.04	.00
Mother used Valium	−.12	−.03	−.10	−.07	−.01	−.02
Mother used antibiotics	−.12	−.08	−.03	−.05	−.07	−.04
Sex (male)	.21	.41	.31	.20	.09	.15
Race (Black)	−.15	−.04	−.03	−.09	.03	−.02
Maternal socioeconomic status	−.12	.01	−.00	−.01	.07	.16
Maternal education	.14	−.01	−.00	.03	−.06	−.15
Maternal marital status	.05	−.02	−.04	.01	−.03	−.06
Maternal welfare status	−.08	−.02	−.02	−.01	.05	.06
Log nicotine	−.26	.00	.04	.04	.05	.10
Log marijuana	−.17	−.03	−.04	−.08	−.01	.07
Log alcohol (prior to pregnancy recognition)	−.15	−.05	.03	.01	.01	.05

Note. Data in this table were tabulated from the maximum number of subjects seen for each exam period, including abstainers (maximum sample size = 553). See Streissguth et al.[30,39] for a full description of measurement of covariates.

Discussion

Dangers of Overreliance on Multiple Regression in Teratology Studies

This paper is the third recent report of alcohol effects on growth (neonatal and postnatal) based on prospective longitudinal studies of prenatal alcohol exposure. Our finding of a significant but transient effect of alcohol upon child's size is in apparent agreement with the conclusions of Greene et al.[13] and in apparent disagreement with those of Day and colleagues.[12,22–24] Both of these other studies use additional measures of alcohol exposure in the third trimester, data that were not gathered in the present study owing to budgetary constraints. However, as the two studies using such third-trimester data disagree with each other, our not having such exposure data is not likely to account for discrepancies be-

tween our findings and theirs. Rather, we would emphasize differences in study design and statistical data analysis between the studies as contributors to the differences in reported results.

The papers by Greene et al. and Day et al. do not report any simple correlations between alcohol dose and outcomes. We therefore cannot tell whether the regressions are "creating" an effect that was not apparent previously or whether they are improving the precision of an estimate of an effect for which there is already evidence. Greene et al. used a backward selection procedure on covariates,[13] whereas Day et al. used forward selection procedures with blocks of variables.[12,22-24] Regression coefficients (and their associated *P* values) computed from hierarchical stepwise procedures can vary greatly over the steps of the procedure in a manner that depends strongly on the joint distribution of all the predictors.[44]

Consider, for example, in Table 2, the unadjusted zero-order correlation of approximately −.05 between 8-month weight and average ounces of alcohol per day prior to pregnancy recognition. A zero-order correlation of −.05 becomes a more negative beta (regression coefficient) when covariates are included that are either positively correlated with both alcohol and 8-month weight or negatively correlated with both alcohol and 8-month weight. Inclusion of an indicator for abstainers confirms a recognized abstainer effect on birthweight, albeit at a lower level, that has been reported by Little,[2] discussed at length by Little and Wendt,[45] and most recently detected again in the EUROMAC study.[46] Linear adjustment principally for one other covariate, log parity, results in a nominally significant regression coefficient for log AAP (partial correlation $\approx -.10, P < .05$). Other highly significant correlates of 8-month weight—sex of child, maternal height, and gestational age—serve only to slightly lower the standard error of the regression coefficient. Inclusion of one additional term, log uses of marijuana, would render the coefficient for alcohol's effect on 8-month weight again insignificant at the nominal .05 level.[47] Thus, although there is evidence of effects of prenatal alcohol exposure on size lasting to 8 months, the reasonably stable significance of the effect of alcohol on birthweight is largely lost.

Our criticisms of the reporting of multiple regression analyses may be summarized in a general problem. The use of multiple-regression methods to search for true effect estimates is much less reliable than their use in searching for optimal predictions (see Mosteller and Tukey[44]). The purposes of covariate analysis in observational studies are two: to adjust for bias and to sharpen the precision of an estimated effect.[48] It is critically important to understand to what extent a regression adjustment appears to present an adjustment for bias and to what extent it simply explains variability in the response and so lowers the estimated variance of a regression slope.

Covariates do not always have enough leverage to substantially improve the precision with which an (alcohol) effect can be estimated, and those that have such leverage, such as socioeconomic status in many studies, ordinarily are not scaled coherently enough to predict the outcome linearly without interaction. That is, in observational studies the attempt to increase the precision of a beta (regression coefficient) inexorably introduces bias as well. As the two purposes of covariate adjustment cannot be disentangled, both must be viewed with suspicion. Multiple regression is safely used primarily as a check that alcohol coefficients are not biased. The disagreement between the conclusions of Greene et al.[13] and those of Day et al.[12,22-24] may serve as a useful caution in this regard: similar analyses of similar outcomes similarly adjusted for similar covariates may result in different findings.

The closer a sampling design can come to a planned experiment, the more stable and reliable the resulting regression slopes. Thus low-dose teratogenic effects can rarely be estimated effectively from samples of convenience, no matter how strenuously one considers covariates. Once a representative sample is enriched by stratified oversampling, as ours was, further oversampling of high doses only increases the precision with which the dose–response relationship can be calibrated; it does not substantially alter the slope of the relationship per se. The longitudinal studies of Greene et al. and Day et al. also oversampled the heavier-drinking ends of their target populations.[13,23,24]

Alcohol Effects on Growth and Comparisons with Functional Outcomes

Our finding that the correlations of alcohol dose with our size measures drop to zero in early childhood is not evidence that the physical effects of prenatal exposure have ceased by that time. We find, for instance, that our composite prenatal dose measure correlates with puberty status at age 14 years. (This will be the topic of another paper.) The effect of alcohol on birthweight we have found here is supported, of course, by studies further oversampling the high-dose range, such as studies of children of alcoholic mothers.[49-51] The absence of a finding for later size is owed to the fact that a broad range of genetic and environmental factors influence size. In contrast, the enduring effects of prenatal alcohol exposure are clear when measured more sensitively in terms of neurobehavioral and/or functional outcomes.[29-34]

Another observation from our analysis pertains to current public health concerns about protocol for fetal alcohol syndrome screening. We noted that of the three children diagnosed with fetal alcohol syndrome, two were small at birth but exhibited typical catch-up growth in weight. The third, although within the normal range of birth size, was very small by age 4, when he was diagnosed. His size deficit persisted thereafter. (This child did not receive a neonatal dysmorphology examination, as his examination was superseded by a birth the same day of a more heavily exposed baby.)

Although a clinical dysmorphology examination detected these offspring with fetal alcohol syndrome, none of their actual birthweights (2840, 2980, and 3140 g) was more than 1 standard deviation below the mean according to usual growth curves, nor would these children technically be classified as having intrauterine growth retardation. Nevertheless, their deviant pattern of increasingly discrepant head circumference up through 8 or 18 months (previously detected in a 10-year follow-up of the original children diagnosed with fetal alcohol syndrome[52]) seems indicative of their underlying central nervous system dysfunction during this period of continuing brain growth.

The discrepant childhood size trajectories of these three children with fetal alcohol syndrome also have important public health implications. The fact that many persons with fetal alcohol syndrome or fetal alcohol effects end up as unusually small adults[53,54] is not helpful in constructing protocols for diagnosis if the onset of this shortfall of size is as variable in time as it was for these three children. Growth deficiency in offspring is one possible consequence of maternal alcoholism. It does not appear to be satisfactory as a unitary screening criterion of fetal alcohol

syndrome, either at birth or at any other age.

Birthweight is far too multifactorial an indicator of generalized fetal suboptimality to serve as a useful outcome for any subtler studies of early developmental processes and their consequences. For instance, earlier research on this same cohort of children demonstrated a substantial association of alcohol dose with a pattern of neurobehavioral deficits at age 7 years.[29-31] But infant size is not a significant covariate of the neuropsychological behaviors we found to be correlated with prenatal alcohol exposure. Thus, even though birth size and 7-year neurobehavioral deficits each appear to show effects of alcohol dose, the birth size deficit is not predictive of the later neurobehavioral deficit. This might be because the aspects of dose (i.e., timing and variability) that are most salient for birth size are different from the binge-oriented measures most salient for later behavioral deficits.[30,31] We are, by and large, not looking at babies of very low birthweight, and the aspects of psychomotor development correlated with low birthweight do not appear to be the aspects that are well correlated with prenatal alcohol exposure. In our sample, low birth size is not a mediator of later alcohol-related problems, nor does it correlate with dose at the extremes of the dose range, nor is it even a predictor of continuing small size. Our study thus strongly suggests that the role of size measures in epidemiologic teratology studies may be overrated.

Conclusions

It is likely that the multifactoriality of birthweight accounts for the inconsistency of alcohol's putative effects on birthweight in the range of studies we reviewed in the Introduction. The effect of alcohol on birthweight, although really there (i.e., it is reliable and stable in stratified samples), can be swamped by other effects—nicotine, race, maternal age, socioeconomic status—some of which are quite a bit more difficult to calibrate, so that the published estimate of the alcohol effect is a function of the range of the covariates and the precise details of the regression technology chosen. Effects of alcohol are much more clearly delineable on outcome measures that are *not* consequences of these other factors of intrauterine growth retardation—outcomes such as aberrant profiles on intelligence tests, speed of information processing, or various subtle but crucial features of vigilance and attention in elementary and junior high school years.

Much of the current public policy debate about the effects of moderate prenatal alcohol exposure has centered on effects as exemplified by birthweight decrements. We believe this emphasis is unfortunate. Although lowered birthweight, with the concomitant risk of intrauterine growth retardation, is clearly of great relevance to public health, its dependence on prenatal alcohol exposure is both quite a bit less sensitive than the dependence of other outcomes we have studied and also much less enduring. The transience of the effect of subextreme alcohol exposure on birthweight appears to be little different in principle from that of nicotine or indicators of mild undernutrition; all these effects vanish over the course of the first few postnatal months (in our analyses). In contrast to effects on size measures, the teratogenic effects of alcohol on certain neurobehavioral sequelae appear to last as long as we have followed our subjects—through 14 years, in the most recent analyses[33,34]—and not only in laboratory testing, but in the classroom as well.[32] □

References

1. Kaminski M, Rumeau-Rouquette C, Schwartz D. Consommation d'alcool chez les femmes enceintes et issue de la grossesse. *Rev Epidemiol Sante Publique.* 1976;24:27–40.
2. Little RE. Moderate alcohol use during pregnancy and decreased infant birth weight. *Am J Public Health.* 1977;67:1154–1156.
3. Kaminski M, Franc M, Lebouvier M, Du Mazaubrun C, Rumeau-Rouquette C. Moderate alcohol use and pregnancy outcome. *Neurobehav Toxicol Teratol.* 1981;3: 173–181.
4. Kuzma JW, Sokol RJ. Maternal drinking behavior and decreased intrauterine growth. *Alcohol Clin Exp Res.* 1982;6:396–402.
5. Olsen J, Rachootin P, Vibeke Schiødt A. Alcohol use, conception time, and birth weight. *J Epidemiol Community Health.* 1983;37:63–65.
6. Wright JT, Waterson EJ, Barrison I, et al. Alcohol consumption, pregnancy and low birthweight. *Lancet.* 1983;1:663–666.
7. Mills JL, Graubard BI, Harley EE, Rhoads GG, Berendes HW. Maternal alcohol consumption and birth weight: how much drinking during pregnancy is safe? *JAMA.* 1984;252:1875–1879.
8. Little RE, Asker RL, Sampson PD, Renwick JH. Fetal growth and moderate drinking in early pregnancy. *Am J Epidemiol.* 1986;123:270–278.
9. Smith IE, Coles CD, Lancaster J, Fernhoff PM, Falek A. The effect of volume and duration of prenatal ethanol exposure on neonatal physical and behavioral development. *Neurobehav Toxicol Teratol.* 1986;8: 375–381.
10. Fried PA, O'Connell CM. A comparison of the effects of prenatal exposure to tobacco, alcohol, cannabis and caffeine on birth size and subsequent growth. *Neurotoxicol Teratol.* 1987;9:79–85.
11. Brooke OG, Anderson HR, Bland JM, Peacock JL, Stewart CM. Effects on birth weight of smoking, alcohol, caffeine, socioeconomic factors, and psychosocial stress. *BMJ.* 1989;298:795–801.
12. Day NL, Jasperse R, Robles N, et al. Prenatal exposure to alcohol: effect on infant growth and morphologic characteristics. *Pediatrics.* 1989;84:536–541.
13. Greene T, Ernhart CB, Sokol R, et al. Prenatal alcohol exposure and preschool physical growth: a longitudinal analysis. *Alcohol Clin Exp Res.* 1991;15:905–913.
14. EUROMAC. A European concerted action: maternal alcohol consumption and its relation to the outcome of pregnancy and child development at 18 months. *Int J Epidemiol.* 1992;21(suppl 1):S1–S87.
15. Hingson R, Alpert JJ, Day N, et al. Effects of maternal drinking and marijuana use on fetal growth and development. *Pediatrics.* 1982;70:539–546.
16. Marbury MC, Linn S, Monson R, Schoenbaum S, Stubblefield PG, Ryan KJ. The association of alcohol consumption with outcome of pregnancy. *Am J Public Health.* 1983;73:1165–1168.
17. Grisso JA, Roman E, Inskip H, Beral V, Donovan J. Alcohol consumption and outcome of pregnancy. *J Epidemiol Community Health.* 1984;38:232–235.
18. Kline J, Stein Z, Hutzler M. Cigarettes, alcohol and marijuana: varying associations with birthweight. *Int J Epidemiol.* 1987;16: 44–50.
19. Walpole I, Zubrick S, Pontre J. Is there a fetal effect with low to moderate alcohol use before or during pregnancy? *J Epidemiol Community Health.* 1990;44:297–301.
20. Rostand A, Kaminski M, Lelong N, et al. Alcohol use in pregnancy, craniofacial features, and fetal growth. *J Epidemiol Community Health.* 1990;44:302–306.
21. Zuckerman BS, Hingson R. Alcohol consumption during pregnancy: a critical review. *Dev Med Child Neurol.* 1986;28:649–654.
22. Day NL, Richardson G, Robles N, et al. Effect of prenatal alcohol exposure on growth and morphology of offspring at 8 months of age. *Pediatrics.* 1990;85:748–752.
23. Day NL, Goldschmidt L, Robles N, et al. Prenatal alcohol exposure and offspring growth at 18 months of age: the predictive validity of two measures of drinking. *Alcohol Clin Exp Res.* 1991;15:914–918.
24. Day NL, Robles N, Richardson G, et al. The effects of prenatal alcohol use on the growth of children at three years of age. *Alcohol Clin Exp Res.* 1991;15:67–71.
25. Russell M, Czarnecki DM, Cowan R, McPherson E, Mudar PJ. Measures of maternal alcohol use as predictors of development in early childhood. *Alcohol Clin Exp Res.* 1991;15:991–1000.
26. Coles CD, Brown RT, Smith IE, Platzman KA, Erickson S, Falek A. Effects of prenatal alcohol exposure at school age, I: physical and cognitive development. *Neurotoxicol Teratol.* 1991;13:357–367.

27. O'Connor MJ, Brill NJ, Sigman M. Alcohol use in primiparous women older than 30 years of age: relation to infant development. *Pediatrics.* 1986;78:444–450.
28. Day NL. Effects of prenatal alcohol exposure. In: Zagon IS, Slotkin TA, eds. *Maternal Substance Abuse and the Developing Nervous System.* San Diego, Calif: Academic Press; 1992:27–43.
29. Streissguth AP, Barr HM, Sampson PD, Bookstein FL, Darby BL. Neurobehavioral effects of prenatal alcohol, part I: research strategy. *Neurotoxicol Teratol.* 1989;11:461–476.
30. Sampson PD, Streissguth AP, Barr HM, Bookstein FL. Neurobehavioral effects of prenatal alcohol, part II: partial least squares analysis. *Neurotoxicol Teratol.* 1989; 11:477–491.
31. Streissguth AP, Bookstein FL, Sampson PD, Barr HM. Neurobehavioral effects of prenatal alcohol, part III: PLS analyses of neuropsychologic tests. *Neurotoxicol Teratol.* 1989;11:493–507.
32. Carmichael Olson H, Sampson PD, Barr HM, Streissguth AP, Bookstein FL. Prenatal exposure to alcohol and school problems in late childhood: a longitudinal prospective study. *Dev Psychopathol.* 1992; 4:341–359.
33. Streissguth AP, Barr HM, Bookstein FL, Sampson PD. *The Enduring Effects of Prenatal Alcohol Exposure on Child Development: Birth through 7 Years.* Ann Arbor, Mich: University of Michigan Press; 1993.
34. Streissguth AP, Sampson PD, Carmichael Olson H, et al. Maternal drinking during pregnancy: attention and short-term memory in 14-year-old offspring: a longitudinal prospective study. *Alcohol Clin Exp Res.* 1994;14:202–218.
35. Riley EP, Vorhees CV. *Handbook of Behavioral Teratology.* New York, NY: Plenum Press; 1986.
36. Streissguth AP, Martin DC, Martin JC, Barr HM. The Seattle longitudinal prospective study on alcohol and pregnancy. *Neurobehav Toxicol Teratol.* 1981;3:223–233.
37. Streissguth AP, Martin DC, Buffington VE. Test–retest reliability of three scales derived from a quantity-frequency-variability assessment of self-reported alcohol consumption. In: Seixas FA, Eggleston S, eds. *Work in Progress in Alcoholism.* Vol 273. New York, NY: Annals of the New York Academy of Sciences; 1976:458–466.
38. Streissguth AP, Barr HM, Martin DC. Offspring effects and pregnancy complications related to self-reported maternal alcohol use. *Dev Pharmacol Ther.* 1982;5:21–32.
39. Cahalan D, Cissin IH, Crossley HM. *American Drinking Practices: A National Study of Drinking Behavior and Attitudes.* New Brunswick, NJ: Rutgers Center of Alcohol Studies Publications; 1969:6. Distributed by College and University Press, New Haven, Conn.
40. Hamill PVV, Drizd TA, Johnson CL, Reed R, Roche AF. NCHS growth curves for children birth–18 years, United States. *Vital Health Stat [11].* 1977; no. 165. DHEW publication PHS 78-1650.
41. Goldstein H. *The Design and Analysis of Longitudinal Studies.* London, England: Academic Press; 1979.
42. Hanson JW, Streissguth AP, Smith DW. The effects of moderate alcohol consumption during pregnancy on fetal growth and morphogenesis. *J Pediatr.* 1978;92:457–460.
43. Graham JM, Hanson JW, Darby BL, Barr HM, Streissguth AP. Independent dysmorphology evaluations at birth and 4 years of age for children exposed to varying amounts of alcohol *in utero. Pediatrics.* 1988;81:772–778.
44. Mosteller F, Tukey JW. *Data Analysis and Regression.* Reading, Mass: Addison-Wesley; 1977.
45. Little RE, Wendt JK. The effects of maternal drinking in the reproductive period: an epidemiologic review. *J Subst Abuse.* 1991;3:187–204.
46. Kaminski M. EUROMAC. A European concerted action: maternal alcohol consumption and its relation to the outcome of pregnancy and child development at 18 months. Chap 10: Relation of findings to other studies. *Int J Epidemiol.* 1992; 21(suppl 1):S79–S81.
47. Barr HM, Streissguth AP, Martin DC, Herman CS. Infant size at 8 months of age: relationship to maternal use of alcohol, nicotine, and caffeine during pregnancy. *Pediatrics.* 1984;74:336–341.
48. Cochran WG. *Planning and Analysis of Observational Studies.* In: Moses LE, Mosteller F, eds. New York, NY: John Wiley & Sons Inc; 1983.
49. Jones KL, Smith DW, Streissguth AP, Myrianthopoulos NC. Outcome in offspring of chronic alcoholic women. *Lancet.* 1974;1:1076–1078.
50. Little RE, Streissguth AP, Barr HM, Herman CS. Decreased birth weight in infants of alcoholic women who abstained during pregnancy. *J Pediatr.* 1980;96:974–977.
51. Aronson M, Olegard R. Children of alcoholic mothers. *Pediatrician.* 1987;14:57–61.
52. Streissguth AP, Clarren SK, Jones KL. Natural history of the fetal alcohol syndrome: a ten-year follow-up of eleven patients. *Lancet.* 1985;2:85–91.
53. Streissguth AP, Aase JM, Clarren SK, Randels SP, LaDue RA, Smith DF. Fetal alcohol syndrome in adolescents and adults. *JAMA.* 1991;265:1961–1967.
54. Lemoine P, Lemoine Ph. Avenir des enfants de mères alcooliques (Étude de 105 cas retrouvés à l'âge adulte) et quelques constatations d'intérêt prophylactique [Outcome in the offspring of alcoholic mothers (study of one hundred and five adults) and considerations with a view to prophylaxis]. *Ann Pédiatr (Paris).* 1992;39: 226–235.
55. Hoaglin DC, Mosteller F, Tukey JW, eds. *Understanding Robust and Exploratory Data Analysis.* New York, NY: John Wiley & Sons Inc; 1983.

Infertility in Women and Moderate Alcohol Use

ABSTRACT

Objective. The purpose of this study was to investigate the relationship between moderate alcohol intake and fertility.

Methods. Interviews were conducted with 3833 women who recently gave birth and 1050 women from seven infertility clinics. The case subjects were categorized based on the infertility specialist's assignment of the most likely cause of infertility: ovulatory factor, tubal disease, cervical factor, endometriosis, or idiopathy. Separate logistic regression models were used to assess the relationship between alcohol use and each type of infertility, adjusted for age, infertility center, cigarette smoking, caffeine use, number of sexual partners, use of an intrauterine device (for tubal disease), and body mass index and exercise (for ovulatory factor).

Results. We found an increase in infertility, due to ovulatory factor or endometriosis, with alcohol use. The odds ratio for ovulatory factor was 1.3 (95% confidence interval [CI] = 1.0, 1.7) for moderate drinkers and 1.6 (95% CI = 1.1, 2.3) for heavier drinkers, compared with nondrinkers. The risk of endometriosis was roughly 50% higher in case subjects with any alcohol intake than in control subjects (OR = 1.6, 95% CI = 1.1, 2.3, at moderate levels; OR = 1.5, 95% CI = 0.8, 2.7, at heavier levels).

Conclusions. Moderate alcohol use may contribute to the risk of specific types of infertility. (*Am J Public Health.* 1994;84:1429–1432)

Francine Grodstein, ScD, Marlene B. Goldman, ScD, and Daniel W. Cramer, MD, ScD

Introduction

Although as many as 20% of married women in the United States have experienced infertility during their lifetime,[1] little attention has been given to the effect that many common exposures, including exposure to alcohol, have on a woman's ability to reproduce.

Much of the research that has been conducted involves studies of animals or alcoholic women. The animal research indicates that decreased steroid hormone levels,[2] reduced ovarian weight,[2,3] and amenorrhea[3] result from alcohol administration to rats or monkeys. Becker et al. reported that alcoholic women experienced higher frequencies of menstrual disturbance, abortion, and miscarriage than control subjects.[4] In a national survey of alcohol use habits and reproductive dysfunction in 917 American women, menstrual problems and gynecologic surgery were increased in women who reported high levels of alcohol intake.[5]

In this study, we examined the effect of moderate alcohol use on fertility in women. We compared self-reported history of alcohol consumption in case subjects who had primary infertility with that of control subjects who recently gave birth.

Methods

Originally, this case–control study was undertaken from 1981 through 1983 to examine the relationship between contraceptive practices and a woman's ability to conceive.[6,7] Case subjects were drawn from women who attended seven infertility clinics in the United States and Canada and who were accepted for work-up and evaluation. Infertility was defined as the inability to conceive after 12 months of unprotected intercourse or the failure to deliver a live-born child.

Case subjects were excluded if their infertility was due to congenital abnormalities, if they were seeking reversal of tubal ligations, or if their husbands had had vasectomies. Among the total number of case subjects who were eligible for interviewing, 5% refused, 10% were not interviewed due to a language barrier or other problem, 10% were lost to follow-up, and 10% had not completed their infertility work-up by the close of the study, leaving 1880 case subjects who were included. Work-ups in these subjects identified the following causes of infertility: ovulatory factor, tubal disease, cervical factor, endometriosis, male factor, or idiopathic infertility. Diagnostic protocols included, when indicated, measurement of basal body temperatures, hormonal studies, endometrial biopsies, postcoital test, hysterosalpingography, and diagnostic laparoscopy.

We assigned each woman to one infertility category based on her most likely cause of infertility, as indicated by the evaluating physician, and designated this cause as the case subject's "first diagnosis." However, infertility is often a multifactorial disorder. In a previous publication, we described in detail the distribution of second diagnoses (accom-

Francine Grodstein and Marlene B. Goldman are with the Department of Epidemiology, Harvard School of Public Health, Boston, Mass. Daniel W. Cramer is with the Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Harvard Medical School, Boston.

Requests for reprints should be sent to Francine Grodstein, ScD, Harvard School of Public Health, Department of Epidemiology, 677 Huntington Ave, Boston, MA 02115.

This paper was accepted November 12, 1993.

TABLE 1—Profile of Case Subjects with Infertility and Control Subjects (in Percentages)

	Control Subjects (n = 3833)	Case Subjects: Ovulatory Factor (n = 431)	Tubal Disease (n = 230)	Cervical Factor (n = 105)	Endometriosis (n = 180)	Idiopathic (n = 104)
Center						
Boston	32.5	30.6	32.2	33.3	26.1	28.8
Vermont	8.3	11.6	21.7	8.6	12.2	15.4
Quebec	21.5	16.2	16.5	20.0	13.3	30.8
Washington, DC	8.4	10.0	10.9	0.9	11.7	12.5
Kentucky	13.1	13.0	3.9	10.5	25.6	1.0
Colorado	16.2	18.6	14.8	26.7	11.1	11.5
Age, y						
<25	40.2	44.3	41.7	33.3	32.2	28.8
25–29	40.1	36.9	36.2	41.0	41.1	38.6
30–34	17.6	14.6	19.1	23.8	25.0	28.8
>34	2.1	4.2	3.0	1.9	1.7	3.8
Education						
High school	29.4	29.5	35.2	29.5	27.2	26.0
College	70.6	70.5	64.8	70.5	72.8	74.0
Cigarette smoking						
Never	53.5	56.1	42.6	42.8	59.4	60.6
Former	16.5	11.4	16.1	14.3	13.9	11.5
Current	30.0	32.5	41.3	42.9	26.7	27.9
Number of sexual partners						
1	45.9	49.9	27.4	38.1	43.4	45.2
2–5	36.7	35.3	42.6	44.8	37.2	36.5
>5	17.4	14.8	30.0	17.1	19.4	18.3
Alcohol intake						
None	63.0	58.0	53.9	56.2	51.7	60.6
Moderate	30.1	33.0	34.4	34.3	38.9	33.7
Heavier	6.9	9.0	11.7	9.5	9.4	5.7

panying disorders that may have contributed to the infertility) among the women in each diagnostic category.[8] The group with ovulatory factor had the lowest percentage of women with an additional disorder (14.2%), whereas the group with a first diagnosis of tubal disease had the highest percentage (26.5%). Idiopathic infertility was always designated as a first and only cause.

Control subjects were recruited from women admitted for delivery of a live birth at hospitals adjacent to the infertility clinics. Adjacent hospitals were used to ensure both that the control subjects would have sought care at the hospital infertility clinic if they had been infertile and that the case subjects would have delivered at the hospital if they had become pregnant. All women admitted for delivery were eligible to be control subjects; among the potential control subjects, 5% refused to be interviewed, 8% had undergone infertility therapy, and 5% were not interviewed because of a language barrier or poor condition of the mother or infant, leaving 4023 control subjects for whom interviews were completed.

This report focuses on the 3833 White control subjects and 1050 White case subjects with no live-born children whose first diagnosis was ovulatory factor, tubal disease, cervical factor, endometriosis, or idiopathic infertility.

Information concerning reproductive history, medical and surgical history, and personal habits was collected by personal interview. Interviews of all case subjects were conducted before their infertility diagnosis. To ensure that all exposure information referred only to the time period before the likely onset of infertility, an "index date" was calculated for each case subject by subtracting the number of months during which she had been trying to conceive from the date on which she had first consulted an infertility specialist. Because the event that distinguished case subjects with primary infertility from control subjects was the first live birth, the index date for each control subject was the estimated date of conception of the first live-born child. If the control subject's pregnancy was planned, the index date was defined as the time at which the couple began attempting conception. Only exposures that occurred before the index date were considered relevant.

During the interview all subjects were asked "what was your average weekly consumption of each of the following beverages: beer (cans per week), wine (glasses per week), and liquor (ounces per week)" before the index date. To determine total alcohol intake per week, 13 g of alcohol was assigned per can of beer, 11 g per glass of wine, and 15 g per ounce of liquor.[9] Women consuming 100 g or less of alcohol per week (approximately one drink or less per day) were considered "moderate" drinkers, whereas those drinking more than 100 g per week were categorized as "heavier" drinkers.

Adjusted odds ratios (ORs) and 95% confidence intervals (CIs) for each level of alcohol intake were calculated by using multiple logistic regression.[10] Separate models were used for each type of infertility. Variables evaluated included age, infertility center, religion, education, body mass index (weight in kilograms/height in meters squared), exercise, cigarette smoking, number of sexual partners, type of contraceptive used, and caffeine intake. Because the study included few non-White women, the analysis was restricted to Whites to control for race as a potential confounder. In all of the final models, age (<25, 25–29, 30–34, >34 years), center (Boston; Vermont; Quebec; Washington, DC; Kentucky; Colorado), cigarette smoking (current, former, never), number of sexual partners (1, 2–5, >5), and caffeine consumption (0–3.0, 3.1–5.0, 5.1–7.0, >7.0 g/week) were included. For analyses of tubal disease, we also controlled for use of an intrauterine device (yes or no); body mass index (<19.0, 19.0–20.9, 21.0–22.9, >23.0 kg/m^2) and exercise (0, 1–6, >6 h/week) were controlled for in models of ovulatory factor.

All tests of significance of the odds ratio were two-tailed. For the analyses involving tests of trend, we categorized alcohol use as explained above, assigned scores to each level of intake, and treated the scored factor as a continuous variable in the logistic regression model. All significant associations were indicated by $P \leq .05$ or a 95% confidence interval excluding 1.

Results

In general, with the exception of women with an ovulatory factor, case subjects tended to be somewhat older than control subjects (Table 1). Educational status was similar in the case and control groups, although those with tubal disease were slightly less educated than control subjects. Similarly, cigarette smoking and sexual habits were comparable in all women, but case subjects with infertility due to tubal disease or cervical factor were more likely to be smokers and to have had multiple sexual partners than were control subjects.

Case subjects who were diagnosed with tubal disease reported the highest levels of alcohol intake. Those with ovulatory factor, cervical factor, or endometriosis also consumed more alcohol than control subjects, although the majority of women in all of the groups reported no alcohol use.

After we adjusted for risk factors, significantly increased odds ratios were observed for infertility due to a first diagnosis of ovulatory factor or endometriosis (Table 2). For ovulatory factor, the odds ratio was 1.3 (95% CI = 1.0, 1.7) for moderate drinkers and 1.6 (95% CI = 1.1, 2.3) for heavier drinkers compared with nondrinkers. This trend of increasing risk of ovulatory infertility with increasing alcohol consumption was statistically significant (test for trend, $P = .005$). There was also a similar trend among women diagnosed with cervical factor (OR = 1.3, 95% CI = 0.8, 2.1 for moderate drinkers; OR = 1.6, 95% CI = 0.8, 3.3 for the heavier drinkers); however, relatively few women with cervical factor reported using alcohol and these estimates were thus unstable. The odds ratio for endometriosis was 1.6 (95% CI = 1.1, 2.3) for moderate drinkers and 1.5 for heavier drinkers (95% CI = 0.8, 2.7) compared with women who did not drink.

Approximately 4% of case subjects with a first diagnosis of an ovulatory factor also had endometriosis, and 12% of those with endometriosis as a first diagnosis were found to have an ovulatory dysfunction. To examine whether the presence of women with both an ovulatory factor and endometriosis had an effect on the risk of infertility observed in these two diagnostic groups, the data were reanalyzed after removing women with both diagnoses (Table 2). The risk of an ovulatory factor associated with alcohol use remained the same (OR = 1.3 for women drinking moderately and OR = 1.6 for those drinking more heavily), whereas the risk increased slightly in women with endometriosis (from an OR of 1.6 to 1.7 at moderate levels and from an OR of 1.5 to 1.8 at higher levels).

TABLE 2—Case Subjects' Risk of Infertility, by Alcohol Use and Type of Infertility

Infertility Type	Moderate Alcohol Use OR[a]	Moderate Alcohol Use 95% CI	Heavier Alcohol Use OR[a]	Heavier Alcohol Use 95% CI
First diagnosis				
Ovulatory factor[b] (n = 431)	1.3	1.0, 1.7	1.6	1.1, 2.3
Tubal disease (n = 230)	1.0	0.7, 1.4	1.2	0.7, 1.9
Cervical factor (n = 105)	1.3	0.8, 2.1	1.6	0.8, 3.3
Endometriosis[c] (n = 180)	1.6	1.1, 2.3	1.5	0.8, 2.7
Idiopathic (n = 104)	0.9	0.5, 1.4	0.7	0.3, 1.6
Ovulatory factor, excluding women with the additional diagnosis of endometriosis (n = 413)	1.3	1.0, 1.7	1.6	1.1, 2.4
Endometriosis, excluding women with the additional diagnosis of ovulatory factor (n = 158)	1.7	1.2, 2.5	1.8	1.0, 3.2

Note. OR = odds ratio; CI = confidence interval.
[a]Adjusted for infertility center, age, number of sexual partners, cigarette smoking, and caffeine intake. For ovulatory factor, hours of vigorous exercise and body mass index were included as confounders; for tubal disease, intrauterine device use was included. Women who reported no alcohol use were used as the reference group.
[b]Test for trend: $P = .005$.
[c]Case subjects who were diagnosed with endometriosis before their infertility evaluation and control subjects who reported a history of endometriosis were excluded.

Discussion

This study demonstrates an association between reported consumption of alcohol and infertility due to ovulatory factor or endometriosis. A small, but significantly increased, risk of ovulatory infertility was observed for women reporting moderate alcohol intake, whereas this risk rose considerably in those women drinking at heavier levels compared with nondrinkers. Increased risks of endometriosis were found at both levels of alcohol intake examined.

Several limitations of these data need to be considered. Because information on exposure history was obtained by self-report, misclassification of alcohol use may have occurred. In particular, approximately 50% to 60% of the women in the case and control groups reported no alcohol intake, slightly more than the 48.3% of women in a national survey of drinking and reproductive dysfunction.[5] To minimize information bias, data were collected during personal interviews that were conducted by trained nurses. Although we were not able to verify the self-reported alcohol use in this population, Willett et al. found that self-report of alcohol intake was highly reproducible in a study of 173 women (Spearman correlation coefficient = 0.90).[11] Our study population was composed exclusively of women of relatively high socioeconomic status, as indicated by the large majority of college-educated participants (approximately 70%). It is possible that their alcohol intake was lower than that of the general population, although to our knowledge no pertinent data have been published. Although this issue may affect the generalizability of our results, it would not affect the validity of our comparison of alcohol intake in case and control subjects, because several measures indicate a similar socioeconomic status in the two groups (education, cigarette smoking, and number of sexual partners were comparable).

This similarity is particularly important in a study such as this one, because the case subjects were recruited from women referred to a specialized medical clinic. Such women may be different, particularly with respect to socioeconomic status, even from the "typical" hospital patient. And alcohol intake could vary with social class. However, as we noted, several socioeconomic indicators are similar among the case and control groups here. In addition, the total household income was similar for case and control subjects, with approximately 50% of both groups falling into the highest income category, suggesting a lack of referral bias in this study.

An additional bias could have resulted from the fact that all the case

subjects were attempting their first pregnancy, whereas some control subjects were multiparous and had to recall exposures before their first pregnancy. In fact, the recall period (i.e., the mean difference between the index date and the interview date) was similar for both groups (approximately 3 years). Although it is possible that subjects' recall of alcohol use 3 years in the past may not have been completely accurate, the recall period was similar for case and control subjects; thus, any potential bias would be nondifferential and lead to an underestimation of the actual risks of infertility associated with alcohol use.[12]

Although the control subjects were instructed to report alcohol use before their first pregnancy, their more recent intake could have been reflected in their report. We reanalyzed the data for the primiparous control subjects only (approximately 50% of the control group) and found risk estimates corresponding to those obtained with the entire control group: for ovulatory factor, the OR was 1.1 (95% CI = 0.9, 1.5) for moderate alcohol use and 1.6 (95% CI = 1.0, 2.6) for heavier alcohol use; for endometriosis, the OR was 1.3 (95% CI = 0.9, 1.9) for moderate intake and 1.4 (95% CI = 0.7, 2.6) for heavier intake.

Still, all the control subjects would have been advised to abstain from alcohol during pregnancy, perhaps leading to an underreport of their alcohol use. If this underreporting were true, we would expect to see increased alcohol intake among all the case subjects. Alcohol was associated with an increased risk of ovulatory factor, cervical factor, and endometriosis; but there was little risk of tubal disease among alcohol users, and a slightly decreased risk of idiopathic infertility was observed.

Because infertility is often a multifactorial disorder, it is also possible that the increased risks of infertility associated with a first diagnosis of ovulatory factor or endometriosis were due to the presence of women with both of these diagnoses. However, when we removed women from each group who were diagnosed with both disorders, a reanalysis of the data did not lead to any appreciable change in the results.

There may be a biological basis for the association that we observed between alcohol use, ovulatory infertility, and endometriosis. Research on the chronic effects of alcohol use on menstruation indicates that alcoholism is associated with early menopause[13] and reduced levels of follicle-stimulating hormone,[13,14] although it remains unknown whether these consequences would also result from the more moderate levels of alcohol use reported in this study. In a cross-sectional study of drinking and reproductive dysfunction, heavy menstrual flow and dysmenorrhea were associated with moderate and high levels of alcohol intake.[5] Both of these factors have been related to an increased risk of endometriosis.[15]

It is possible that women with symptoms of endometriosis (specifically, pain in the pelvic area) drink alcohol to help alleviate their pain. However, when we examined use of alcohol by the case subjects with endometriosis, we found no significant difference between women reporting none to mild dysmenorrhea and those reporting moderate to severe dysmenorrhea. Approximately 50% of case subjects in both those groups were alcohol users.

There is presently little information available on the specific effects of moderate alcohol intake on reproductive capacity. This is one of few investigations that has examined the consequences of moderate alcohol use on distinctive types of fertility problems. We found the largest effect of alcohol use to be on the hormonally associated fertility disorders. It is of interest that an increased risk of breast cancer has also been associated with moderate alcohol use.[16,17] No mechanism has been established yet to explain this association, but breast cancer is most likely an estrogen-dependent neoplasm,[18] and both ovulatory function and endometriosis are related to estrogen levels. □

Acknowledgments

This research was supported by contract N01-HD-02822 and grant R29-HD-23718 from the National Institute of Child Health and Human Development and by Reproductive Hazards in the Workplace, Home, Community and Environment Research Grant 15-138 from the March of Dimes Birth Defects Foundation. Dr Grodstein was supported by a National Institute of Environmental Health Sciences National Research Service Award (5T32 ES07069) through the Harvard School of Public Health.

We are indebted to Louise Ryan, PhD, for her biostatistical advice and to the collaborators at each center: Bruce Albrecht, MD, University of Colorado Health Sciences Center, Denver; Serge Belisle, MD, Department of Obstetrics and Gynecology, Centre Hospitalier, Universite de Sherbrooke, Sherbrooke, Quebec; Mark Gibson, MD, Department of Obstetrics and Gynecology, University of Vermont, Burlington; Isaac Schiff, MD, formerly with the Department of Obstetrics and Gynecology, Brigham and Women's Hospital, and currently with Massachusetts General Hospital, Boston; Robert Stillman, MD, Department of Obstetrics and Gynecology, George Washington University Medical Center, Washington, DC; Irwin Thompson, MD, New England Fertility and Gynecology Associates, Department of Obstetrics and Gynecology and Charles A. Dana Research Center, Beth Israel Hospital, Boston; and Emery Wilson, MD, Department of Obstetrics and Gynecology, University of Kentucky Medical Center, Lexington.

References

1. *Infertility: Medical and Social Choices.* Washington, DC: Office of Technology Assessment; 1988.
2. Van Thiel DH, Gavaler JS, Lester R. Alcohol-induced ovarian failure in the rat. *J Clin Invest.* 1978;61:624–632.
3. Mello NK, Bree MP, Mendelson JH, Ellingboe J. Alcohol self-administration disrupts reproductive function in female macaque monkeys. *Science.* 1983;221:577–679.
4. Becker U, Tonnesen H, Kaas-Claesson N, Gluud C. Menstrual disturbances and fertility in chronic alcoholic women. *Drug Alcohol Depend.* 1989;24:75–82.
5. Wilsnack SC, Klassen AD, Wilsnack RW. Drinking and reproductive dysfunction among women in a 1981 national survey. *Alcohol Clin Exp Res.* 1984;8:451–458.
6. Cramer DW, Schiff I, Schoenbaum SC, et al. Tubal infertility and the intrauterine device. *N Engl J Med.* 1985;312:941–947.
7. Cramer DW, Goldman MB, Schiff I, et al. The relationship of tubal infertility to barrier method and oral contraceptive use. *JAMA.* 1987;257:2446–2450.
8. Grodstein F, Goldman MB, Ryan L, Cramer DW. Female infertility and caffeinated beverages. *Am J Epidemiol.* 1993;137:1353–1360.
9. *Agricultural Handbook.* Washington, DC: US Department of Agriculture; 1963–1988. Series No. 8.
10. Breslow NE, Day NE. *Statistical Methods in Cancer Research: The Analysis of Case-Control Studies.* Lyon, France: International Agency for Research on Cancer; 1980;1:192–246.
11. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol.* 1985;122:51–65.
12. Rothman KJ. *Modern Epidemiology.* Boston, Mass: Little Brown & Co Inc; 1986:84–89.
13. Gavaler JS, Van Thiel DH. Reproductive consequences of alcohol abuse: males and females compared and contrasted. *Mutat Res.* 1987;186:269–277.
14. Hadi HA, Hill JA, Castillo RA. Alcohol and reproductive function: a review. *Obstet Gynecol Surv.* 1987;42:69–74.
15. Cramer DW, Wilson E, Stillman RJ, et al. The relation of endometriosis to menstrual characteristics, smoking, and exercise. *JAMA.* 1986;255:1904–1908.
16. Longnecker MP, Berlin JA, Orza JM, Chalmers TC. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA.* 1988;260:652–656.
17. Willett WC, Stampfer MJ, Colditz GA, Rosner BA, Hennekens CH, Speizer FE. Moderate alcohol consumption and risk of breast cancer. *N Engl J Med.* 1987;316:1174–1180.
18. Kelsey JL. A review of the epidemiology of human breast cancer. *Epidemiol Rev.* 1979;1:74–109.

Trends in the US Prevalence of Drug-Using Parturient Women and Drug-Affected Newborns, 1979 through 1990

Marvin Dicker, PhD, and Eldin A. Leighton, PhD

ABSTRACT

Objectives. There has been a lack of reliable national data on the number of pregnant women using drugs and the number of newborns affected by such use. The major reasons for this lack have been inadequate sampling and data collection procedures and the lack of a risk assessment perspective in analysis. This paper corrects for these inadequacies.

Methods. Data from 1979 through 1990 from the National Hospital Discharge Survey, an annual survey by the National Center for Health Statistics, were analyzed.

Results. Between 1979 and 1990 there was a 576% increase in the rate of discharges of drug-using parturient women in the United States and a 456% increase in the rate of discharges of drug-affected newborns. After adjustment for underreporting, a "best estimate" of the number of discharges from 1988 through 1990 was about 88 000 per year for drug-using parturient women and about 48 000 per year for drug-affected newborns.

Conclusions. Although the data support the occurrence of a national epidemic of drug use among pregnant women during the 1980s, the size and severity of this epidemic have been overstated. (*Am J Public Health.* 1994;84:1433–1438)

Introduction

Although much has been written about an epidemic of drug use among pregnant women in the United States during the 1980s, two recent research papers have noted that there have been almost no reliable national data on the prevalence or incidence of such use in particular years or on the nature of the historical trend associated with such use.[1,2] National data on the use of drugs by pregnant women is of special interest because of the effect drug use might have on the health of women and their newborns. The lack of reliable national data has made it difficult to estimate both the extent and seriousness of the national problem and the amount of medical and other societal resources needed to deal with the problem.

This lack of national data was partially corrected in a 1991 paper in which we used data from the National Hospital Discharge Survey (NHDS) to examine the historical trend in diagnosed drug-related births in US hospitals between 1979 and 1987.[3] Of all the recent studies of drug use affecting newborns, this is the only study with both a representative national sample and reliable data collection techniques. This study found a 339% increase between 1979 and 1987 in the estimated discharge rate for newborns with a drug diagnosis. However, after adjustment for underreporting, the absolute number of newborns with a drug diagnosis in 1987 was estimated to be about 38 000—an estimate that was much smaller than others reported in the literature. In this paper we expand the previous study by presenting data extending the historical trend for drug-affected newborns to 1990.

We also present data on the historical trend between the years 1979 and 1990 for the US population of pregnant women diagnosed as using drugs at the time they entered the hospital for delivery. These women will be labeled "drug-using parturient women" to distinguish them from the larger population of pregnant women who used drugs at some time during their pregnancy.[2] The data on parturient women are from the same set of annual surveys used for our previous study of drug-affected newborns;[3] however, the data on newborns and the data on parturient women were collected independently of one another within each annual survey. This means that information about a mother cannot be linked to information about her child. As a result, each annual estimate of the number and rate of drug-using parturient women is an independent estimate of the number and rate of drug-exposed newborns.

Methods

Data Sources

Data for the present study are from the NHDS, which is conducted annually in the United States by the National Center for Health Statistics (NCHS). Survey responses were analyzed from public use files for the 12 years 1979 through 1990. This survey uses a stratified probability sample and weights to pro-

At the time of the study, the authors were with the Washington Consulting Group, Washington, DC. Marvin Dicker was also with the National Institute on Drug Abuse, Rockville, Md.

Requests for reprints should be sent to Marvin Dicker, PhD, 10273 Green Holly Terrace, Silver Spring, MD 20902.

This paper was accepted November 18, 1993.

Note. The views expressed in this paper are the authors' and do not necessarily represent those of the Washington Consulting Group or the National Institute on Drug Abuse.

TABLE 1—Diagnostic Codes Used in Study

Description of Diagnosis	ICD-9-CM Code(s)[13]
General	
Drug psychoses	All 292's
Drug dependence	All 304's
Nondependent abuse of drugs	305.2–305.9
Polyneuropathy due to drugs	357.6
Pregnancy and birth	
Drug dependence complicating pregnancy	648.3
Suspected damage to fetus from drugs	655.5
Narcotics/hallucinogens affecting fetus/newborn	760.72, 760.73
Drug withdrawal syndrome in newborn	779.5
Poisoning	
Poisoning by analgesics	965.0, 965.8, 965.9
Poisoning by sedatives and hypnotics	967.0–967.6, 967.8, 967.9
Poisoning by other central nervous system depressants	968.0, 968.5
Poisoning by psychotropic agents	969.0–969.9
Poisoning by central nervous system stimulants	970.0–970.9
Adverse effects of drugs properly used in therapy	E935.0–E935.2, E937–E940

duce national estimates of all discharges from the approximately 6000 nonfederal short-stay hospitals in the United States. As these hospitals account for more than 99% of all births occurring in the United States,[4,5] very few births are not represented in the sample. Overall details of the NHDS's data collection, sampling, and coding strategies can be found in annual reports published by NCHS[6] and in a recent paper by Shimizu.[7]

NCHS redesigned the sample in 1988. Therefore, differences in NHDS estimates between 1987 and 1988 could result from sampling changes rather than changes in hospital utilization. In this study, however, we found no statistically significant differences between the years 1987 and 1988, so the sampling change does not present a problem when interpreting the overall trend presented here.

As stated above, the NHDS uses a complex, stratified sampling scheme. Therefore, to avoid confusion, all national data in the text and tables of this paper attributed to the NHDS are weighted population estimates unless otherwise stated. The number of sample cases underlying each drug-related estimate is found in the footnotes to the tables.

Statistical Techniques

NCHS has established reliability guidelines for using survey results from the NHDS.[6,8,9] To comply with these guidelines, sample cases in this study were aggregated over several years when necessary to create groups of sample size greater than or equal to 30 for newborns and greater than or equal to 60 for parturient women. However, because of the sample redesign referred to above, data for the years before 1988 were never combined with data for 1988 and later years. For combined time periods, the estimate given is an annual average estimate for each year included in the time period.[6,8–11]

Research on the incidence of drug use during pregnancy has focused on the number of cases rather than on rates. The number of cases allows for estimates of the costs or amount of resources needed to deal with problems associated with drug use during pregnancy. However, rates are more useful for assessing such relationships as changes over time, differences between populations, and differences between studies, because rates standardize for differences in population size. Here we present estimates of both numbers and rates. All differences between estimates discussed in this paper are statistically significant ($P < .05$) as determined by a multiple t test based on the Bonferroni inequality.[12]

Identification of Drug Users

Table 1 lists the diagnostic categories considered to indicate a drug-affected newborn or drug use by a parturient woman. A discharge was coded as drug-related if one or more of these *International Classification of Diseases*, Ninth Revision, Clinical Modification[13] (ICD-9-CM) diagnoses was found on the discharge record. However, not all the diagnostic categories listed were actually found.

The types of ICD-9-CM codes selected to identify drug-related discharges can affect both the shape of the historical trend and the total number of cases found. This study follows Rice and Kelman in including E-codes in the set of identifying codes.[10] E-codes indicate adverse effects associated with drugs used properly in treatment. They are included here on the assumption that some physicians may have substituted E-codes for other, more stigmatizing, drug codes when recording drug diagnoses. As this inclusion expands the universe for identifying drug discharges, this paper presents a worst-case scenario in its estimates of the total number of drug-related discharges at any given time.

Results

Drug Use among Parturient Women

Table 2 shows the historical trend in the United States in the estimated average annual number and rate of discharges for drug-using parturient women for the whole decade of the 1980s. The results shown in this table indicate that there was a continuing increase in these numbers and rates during this time. For example, the average annual number of discharges involving drug-using parturient women increased by 624%, from 3799 discharges in the 1979 through 1981 time period (95% confidence interval [CI] = 2389, 5209) to 27 506 discharges in 1990 (95% CI = 23 062, 31 950). The average annual discharge rate also showed a large increase (576%) over the 12 years examined, rising from 10.1 discharges per 10 000 parturient women in the 1979 through 1981 time period to 68.3 in 1990.

Although the impression given by Table 2 is of an alarming continuous increase in the number and rate of discharges for drug-using parturient women during the 1980s, this rate of increase was continuously diminishing. The rate of increase in the number of discharges for drug-using parturient women between the 1979 through 1981 and 1982 through 1984 time periods was 102% (a doubling of the number). This rate of increase dropped continuously throughout the decade until, finally, between 1989 and 1990, the number of discharges for drug-using parturient women decreased by 13%. These statistics demonstrate that drug use by parturient women was increasing at a slower rate throughout the decade, so that by the 1988 through 1990 time period a leveling

TABLE 2—Estimated Number and Rate of Discharges of Parturient Women with a Drug Diagnosis from Nonfederal Short-Stay Hospitals: United States, 1979 through 1990

Year of Discharge	Average Annual No. of Parturient Women		SE for Drug Discharges	Estimated Range of the No. of Parturient Women with a Drug Diagnosis		Rate per 10 000 Parturient Women	Increase in Annual Average No. from Previous Time Period, %
	Total	With a Drug Diagnosis		Low	High		
1979–1981	3 773 719	3 799	719	2 389	5 209	10.1	NA
1982–1984	3 924 429	7 664	870	5 959	9 369	19.5	102
1985–1986	3 808 084	14 536	1 199	12 186	16 886	38.2	90
1987	3 910 926	19 597	1 703	16 259	22 935	50.1	35
1988	3 780 654	28 375	3 567	21 385	35 365	75.1	45
1989	3 936 703	31 706	4 255	23 367	40 045	80.5	12
1990	4 025 456	27 506	2 267	23 062	31 950	68.3	-13

Note. Years were combined to meet National Center for Health Statistics (NCHS) reliability guidelines. Sample sizes for parturient women with a drug diagnosis were as follows: for 1979–1981, n = 68; for 1982–1984, n = 121; for 1985–1986, n = 131; for 1987, n = 98; for 1988, n = 237; for 1989, n = 315; for 1990, n = 247. A discharge was counted as a drug-diagnosis discharge if any of up to seven diagnoses on the discharge record was a drug diagnosis. Standard errors were estimated as follows: before 1988, from charts in the National Hospital Discharge Survey (NHDS) annual summaries[6]; after 1988, from empirical equations in the documentation accompanying the public use tapes. The "low" and "high" estimates are based on 95% confidence intervals.

Source. National Hospital Discharge Survey, NCHS (original tabulations from public use tapes).

off had occurred in the number and rate of discharges.

A Comparison of Parturient Women and Newborns

The historical trend. The number of drug-using parturient women may be considered a surrogate measure for the number of newborns exposed to drugs at the time of delivery. But measures of exposure do not necessarily indicate the number of exposed persons with detrimental effects. The toxicity and risk assessment literatures stress the importance of maintaining a distinction, both theoretically and empirically, between the number of persons exposed and the number of persons who have received a biologically effective dose.[14–18] Therefore, a comparison between the data on drug-using parturient women and the data on drug-affected newborns may shed some light on the relationship between exposure to drugs (in the aggregate) at birth and aggregate adverse health outcomes associated with that exposure.

The basis for such a comparison was published in the paper on newborns discussed above.[3] In that paper we examined the historical trend between 1979 and 1987 in the number of newborns discharged with a drug diagnosis. All newborns with one or more drug diagnoses on their discharge records were labeled as "drug-affected." While the phrase "drug-affected newborn" does not imply any adverse health condition other than the existence of a drug diagnosis on the discharge record, many newborns with a drug diagnosis had adverse health problems at birth. For example, about 84% of all newborns with a drug diagnosis had either drug withdrawal symptoms (ICD-9-CM code 779.5) or a nondrug medical diagnosis indicating health problems other than drug exposure. The remaining 16% were coded 760.72, indicating the presence of narcotics without any specified adverse health outcome.

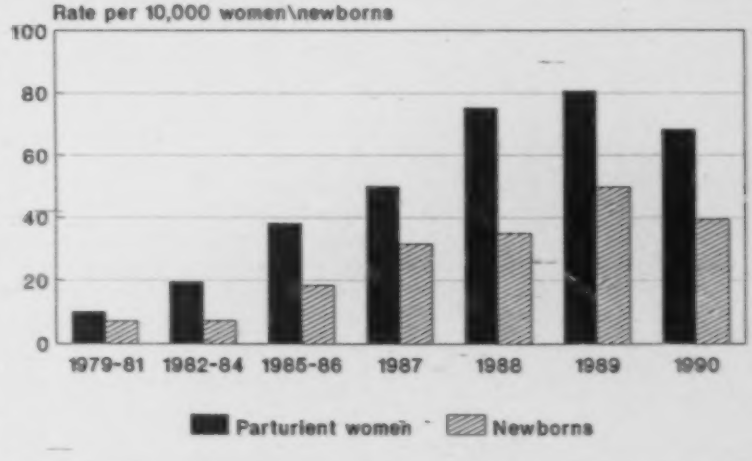


FIGURE 1—Average annual rates of women and newborns with drug-related diagnoses: United States, 1979 through 1990.

Combining data on drug-affected newborns from the previous paper with new data for the 1988 through 1990 time period calculated for this paper, Figure 1 compares the historical trend between 1979 and 1990 in the discharge rate for drug-using parturient women and drug-affected newborns (as defined above). The overall curve for drug-using parturient women is very similar to the overall curve found for drug-affected newborns. Both show a rapid and continuous rise in the rate of drug-related discharges until 1989 and then a decrease in 1990. The overall similarity in the shape of the historical trend for the two populations suggests two things: first, that these two populations represent two aspects of one phenomenon, drug exposure among newborns, and second, that each of the

TABLE 3—Estimated Average Annual Number and Rate of Discharges with a Drug Diagnosis for Newborns and Parturient Women, for the Combined 3-Year Period 1988 through 1990

Population	Average Annual No. of Discharges: Total	Average Annual No. of Discharges: With a Drug Diagnosis	SE for Drug Discharges	Estimated Range of the No. of Discharges with a Drug Diagnosis: Low	Estimated Range of the No. of Discharges with a Drug Diagnosis: High	Rate per 10 000
Newborns	3 828 702	15 846	2 519	10 908	20 784	41.4
Parturient women	3 914 271	29 196	3 334	22 660	35 731	74.6

Note. Years were combined to get a best estimate. Sample size for parturient women with a drug diagnosis for the combined years is 799. Sample size for drug-affected newborns for the combined years is 512. A discharge was counted as a drug-diagnosis discharge if any of up to seven diagnoses on the discharge record was a drug diagnosis. Standard errors were estimated from empirical equations in the documentation accompanying public use tapes for these years. The "low" and "high" estimates are based on 95% confidence intervals.

Source. National Hospital Discharge Survey, NCHS (original tabulations from public use tapes).

populations is a reliable indicator of the national historical trend in drug-related discharges associated with pregnancy in short-stay hospitals in the United States. The decrease in the rate for 1990 is suggestive because it is found in both populations examined. Whether this 1-year decline in the otherwise continuous increase represents the beginning of a long-term decrease or is merely a short-term phenomenon cannot be ascertained at this time.

Although the curves are similar, the rates shown in Figure 1 also suggest that the two populations do not represent exactly the same phenomenon. In every time period examined, the rate per 10 000 for drug-using parturient women was notably higher than the rate per 10 000 for drug-affected newborns. Of the seven possible comparisons in Figure 1, five differences are statistically significant at the .05 level, and these differences are important in an absolute sense. On the average, over all 12 years, the number of drug-affected newborns was only 53% of the number of parturient women diagnosed as using drugs.

Best estimates for 1988 through 1990. For comparisons with other studies in the literature, average estimates for the 3 years 1988 through 1990 are desirable because the single estimates in this paper for these years are not statistically different from one another ($P > .05$). Furthermore, the direction of the historical trend is inconsistent, first rising and then falling over this time period. This seems to indicate a leveling of the historical trend, which is best represented by a single average estimate. Such an estimate is also based on a larger number of sample cases, thus increasing statistical reliability.

Table 3 indicates that the average annual estimate of the number of discharges for drug-using parturient women in the United States during 1988 through 1990 was 29 196 discharges per year (95% CI = 22 668, 35 731). For drug-affected newborns, the average annual estimate for the same 3 years was 15 846 discharges per year (95% CI = 10 908, 20 784). The associated average annual rate of drug use per 10 000 parturient women was 74.6; the average annual rate of newborns who were born drug-affected was 41.4 per 10 000.

Discussion

The Historical Trend

The results reported here give strong support to the perception that an epidemic of drug use among pregnant women occurred in the United States between 1979 and 1990,[19] although changes in physicians' reporting practices and increased use of improved drug testing technology may have also contributed to the results.[3]

A reverse trend for drug use has been reported among women of childbearing age (i.e., between the ages of 15 and 45).[1] Khalsa and Gfroerer, using the Household Survey on Drug Abuse sponsored by the National Institute on Drug Abuse, show that the rate of previous-month drug use among these women decreased from about 1500 drug-using women per 10 000 in 1985 to about 800 in 1990.[1] This is a 47% drop in the rate for previous-month users over this 5-year period.

These contrasting results suggest that modeling drug use among pregnant women on the basis of drug use among women of childbearing age, as was done in a recent Institute of Medicine study,[20] would lead to erroneous estimates of both direction in historical trend and incidence of use. Drug-using parturient women seem to be a distinct subpopulation within the larger population of childbearing women, who have historically been immune to the social forces causing drug use to decrease. Instead, countervailing forces increased drug use among this subpopulation.

Assessing Incidence

Each annual discharge in the NHDS represents both a new case (incidence) and a specific person because it is rare for a woman to deliver more than one child in a 12-month period. Given that this assumption is valid, neither the data for parturient women nor the data for newborns support the higher national estimates of the number of drug-using pregnant women (or drug-exposed newborns) presented in the literature for the years 1988 through 1990. Most of these estimates have ranged between 100 000 and 739 000 women or newborns with a drug diagnosis.[20–26] The best annual estimate for these years based on the NHDS is only 29 196 drug-using women or newborns exposed to drugs. The best annual estimate of the number of drug-affected newborns is 15 846. These numbers raise questions about the reliability and validity of the higher national estimates found in the literature. Because none of the other estimates were derived from surveys with both representative samples of all births in the United States and reliable data collection techniques, these findings suggest that the number of drug-using pregnant women and drug-affected newborns was much smaller in this time period than previously thought.

Adjusting for Underreporting

One criticism of using the NHDS to study the incidence of drug-related discharges is that the survey produces low numbers because drug use is underreported on discharges. We used findings on underreporting from local drug testing studies comparing self-reports with biomarker indications (urine, hair, meconium, and so forth) to construct an algorithm for adjusting estimates from the

NHDS.[3] While these studies varied greatly in the amount of false reporting found, an upper bound for underreporting could be assessed. The results suggested that findings from the NHDS should be weighted by a multiplier of 3 for estimates from 1986 through 1990.[3] This statistical adjustment for underreporting is found in Table 4.

Because the adjusted numbers in Table 4 represent statistical adjustments based on a limited number of studies from local hospitals in areas of high drug use, they should be thought of as provisional, but they are nevertheless revealing. They indicate that after adjustment for probable underreporting, only about 88 000 pregnant women per year can be estimated to have used illicit drugs at the time of delivery in the United States during the 1988 through 1990 time period (95% CI = 67 984, 107 192). During the same time period, after adjustment for underreporting, the estimate is that only about 48 000 newborns might have been adversely affected by drugs annually in the United States (95% CI = 32 726, 62 350). The equivalent estimated national rates per 10 000 are 224 drug-using parturient women and 124 drug-affected newborns.

TABLE 4—Estimated Average Annual Number and Rate of Discharges with a Drug Diagnosis for Newborns and Parturient Women, Adjusted for Underreporting, for the Combined 3-Year Period 1988 through 1990

Population	Average Annual No. with a Drug Diagnosis	Adjustment Factor	Adjusted No. with a Drug Diagnosis	Adjusted SE	Adjusted Range of the No. with a Drug Diagnosis		Adjusted Rate per 10 000
					Low	High	
Newborns	15 846	3	47 538	7 557	32 726	62 350	124.2
Parturient women	29 196	3	87 588	10 002	67 984	107 192	223.8

Assessing Risks to the Newborn

Perhaps the most interesting finding of this study is the gap in incidence between the number of drug-using parturient women and the number of drug-affected newborns. This gap was found in each time period and was always in the same direction. While the gap varied in size among time periods, on average over all 12 years the number of discharges for drug-affected newborns was estimated to be only about half (53%) of the number of discharges for drug-using parturient women.

Two possible explanations (or hypotheses) for this phenomenon come to mind, neither of which can be tested with the data on hand. The first, and most likely, explanation is that the gap represents an exposure–toxicity relationship. In the exposure–toxicity hypothesis, the gap represents the aggregate risk that a drug-exposed newborn will have an adverse health outcome. There is support for this explanation in the literature. For example, the biomarker literature suggests that many exposed newborns appear normal at birth despite a finding of drug use by the mother.[27] Also, Chasnoff reports that fewer than half of newborns exposed to cocaine in utero show any adverse effects at the time of birth.[22] This explanation of the gap is also consistent with the general model for environmental risk assessment research recommended by the Environmental Protection Agency.[15,16]

The second explanation is that the gap represents reporting errors. For the reporting error hypothesis to be likely, the recording of diagnoses for parturient women and newborns would have to represent independent acts wherein knowledge of the medical condition of one partner in the mother–newborn pair does not affect the diagnosis given the other partner. This condition does not hold for the diagnoses under investigation here. For example, it is hard to believe that attending medical personnel who are aware that the mother used drugs prior to delivery would not diagnose the newborn as drug-affected unless the newborn appeared unaffected by the mother's use of drugs. This implies the hypothesis that, during the 1980s, it was customary for medical personnel to give a known drug-exposed newborn a drug diagnosis only when the newborn either had a diagnosable adverse medical condition or tested positive for drugs. From a risk assessment perspective, therefore, a parturient woman could be correctly diagnosed as drug-using while her newborn was not diagnosed as drug-affected, implying an exposure–toxicity relationship of less than one.

Errors in the opposite direction also do not invalidate the exposure–toxicity hypothesis. These errors occur when a physician discovers that a newborn has tested positive for drugs but fails to classify the mother as drug-using. Such errors underestimate the number of drug-using parturient women, making the incidence gap smaller than it otherwise would be.

The consistency, size, and direction of the gap, plus the lack of independence in diagnoses, make it unlikely that the gap is primarily the result of reporting errors. Rather, the gap represents the risk, when examining aggregates, of an adverse health outcome among newborns exposed to drugs near the time of delivery.

While this risk (approximately 0.5) suggests a very high rate of toxicity if all cases are attributed to drug exposure, it raises questions about some methods used in the literature to estimate the number of drug-affected newborns at birth. For example, studies that use the number of drug-using pregnant women as a surrogate measure of the number of newborns adversely affected by drugs would overstate this number by about 100%. The key variable in assessing the burden on hospital services of exposure to a toxic substance is not the number exposed but the number adversely affected by the exposure. If there is no adverse health outcome, then there is no medical reason for special treatment, longer lengths of stay, or other extra services that lead to higher costs and additional burdens on a hospital system. Studies that use an estimated national incidence of more than 50 000 adversely affected newborns to estimate the national hospital costs associated with drug-related births (such as that of Phibbs et al.[28]) overstate the burden on the US hospital system.

Furthermore, the results presented here on the number of drug-affected newborns only indicate an upper bound for the possible toxic effect of drug use (in the aggregate) during pregnancy. This is because evidence is accumulating that many pathological conditions in the newborn thought to be directly attributable to drugs may instead be attributable to other causes associated with drug use by the mother.[2,14,29–32]

Our findings support the perception that drug use among pregnant women steadily increased during the 1980s until 1988, when it appeared to level off. However, the findings do not support

other perceptions concerning drug use during pregnancy. The estimated number of drug-using pregnant women and drug-affected newborns in the 1988 through 1990 time period was much smaller than would be expected on the basis of other reports in the research literature and popular media. The reason for the discrepancy seems to be a lack of representative samples and reliable data collection methods in the studies from which these reports were derived. Finally, the risk of an adverse health outcome among newborns exposed to drugs was found to be much smaller than the risk implied in many discussions of drug use during pregnancy. Above all, our findings do not support the model that one instance of drug exposure in the mother always yields an adverse health outcome in the newborn. One must conclude, therefore, that while an epidemic of drug use among pregnant women did take place during the 1980s, both the size and the severity of the epidemic have been overstated. □

Acknowledgments

This research was initiated at the National Institute on Drug Abuse (NIDA) by Marvin Dicker while he was an employee of that agency. He wishes to acknowledge the general support of NIDA at that time. In particular, he wishes to acknowledge the support of Loretta P. Finnegan, MD. Eldin A. Leighton's participation in the research was partially supported by the National Institute on Drug Abuse under contract no. 271-91-8324.

References

1. Khalsa JH, Gfroerer J. Epidemiology and health consequences of drug abuse among pregnant women. *Semin Perinatol.* 1991;15: 265–270.
2. Mayes LC, Granger RH, Bornstein MH, Zuckerman B. The problem of prenatal cocaine exposure: a rush to judgement. *JAMA.* 1992;267:406–408.
3. Dicker M, Leighton EA. Trends in diagnosed drug problems among newborns: United States, 1979–1987. *Drug Alcohol Depend.* 1991;28:151–165.
4. *Hospital Statistics, 1987 Edition.* Chicago, Ill: American Hospital Association; 1987.
5. National Center for Health Statistics. Advance report of final natality statistics, 1988. *Month Vital Stat Rep.* 1990;39(4): suppl.
6. National Center for Health Statistics. National Hospital Discharge Survey annual summaries. *Vital Health Stat [13].* 1979–1987; nos. 60, 64, 72, 78, 83, 84, 91, 96, 99.
7. Shimizu IM. The new statistical design of the National Hospital Discharge Survey. In: *1990 Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association; 1990: 702–707.
8. Pokras R. *Statistical Note (On Calculating Standard Errors When Combining Years of Data from the National Hospital Discharge Survey).* Hyattsville, Md: National Center for Health Statistics; February 10, 1988.
9. Shimizu IM. *Rel-Variances for NHDS Estimates Based on 2 Years of Data.* Hyattsville, Md: National Center for Health Statistics; August 5, 1982.
10. Rice DP, Kelman S. Measuring comorbidity and overlap in the hospitalization cost for alcohol and drug abuse and mental illness. *Inquiry.* 1989;26:249–260.
11. Gfroerer JC, Adams EH, Moien M. Drug abuse discharges from non-federal short-stay hospitals. *Am J Public Health.* 1988;78: 1559–1562.
12. Levy PS, Lemeshow S. *Sampling for Health Professionals.* Belmont, Calif: Lifetime Learning Publications; 1980.
13. *International Classification of Diseases,* Ninth Revision, Clinical Modification. Hyattsville, Md: National Center for Health Statistics; 1980. DHHS publication PHS 80-1260.
14. Neuspiel DR, Hamel SC. Cocaine and infant behavior. *J Dev Behav Pediatr.* 1991;12:55–64.
15. *Guidelines for Estimating Exposure.* Washington, DC: US Environmental Protection Agency; September 1986.
16. *Guidelines for Developmental Toxicity Risk Assessment.* Washington, DC: US Environmental Protection Agency; December 1991.
17. Goldstein BD. The scientific basis for policy decisions. In: Gordis L, Libauer CH, eds. *Epidemiology and Health Risk Assessment.* New York, NY: Oxford University Press; 1988:11–17.
18. Rowland Hogue CJ, Brewster MA. Developmental risks: advances in health assessment. In: Gordis L, Libauer CH, eds. *Epidemiology and Health Risk Assessment.* New York, NY: Oxford University Press; 1988:61–81.
19. Mausner JS, Kramer S. *Mausner and Bahn Epidemiology.* 2nd ed. Philadelphia, Pa: W.B. Saunders Co; 1985.
20. Gerstein DR, Harwood HJ, eds. *Treating Drug Problems,* Vol. I. Washington, DC: National Academy Press; 1990:85.
21. Chasnoff IJ. Drug use and women: establishing a standard of care. In: Hutchings DE, ed. Prenatal Abuse of Licit and Illicit Drugs. *Ann N Y Acad Sci.* 1989;562:208–210.
22. Chasnoff IJ. Drugs, alcohol, pregnancy, and the neonate: pay now or pay later. *JAMA.* 1991;266:1567–1568.
23. *Crack Babies.* A report of the Office of Inspector General, Office of Evaluation and Inspections. Washington, DC: US Dept of Health and Human Services; June 1990. OEI-03-89-01540.
24. *Drug Exposed Infants, a Generation at Risk.* A report to the Chairman, Committee on Finance, United States Senate. Gaithersburg, Md: US General Accounting Office; June 1990. GAO/HRD-90-138.
25. *National Drug Control Strategy.* Washington, DC: Executive Office of the President, Office of National Drug Control Policy; January 1990. Available from Superintendent of Documents, Washington DC 20402-9325. Order No. S/N 040-000-00543-9.
26. Finnegan LP, Kandall SR. Maternal and neonatal effects of alcohol and drugs. In: Lowinson JH, Ruiz P, Millman RB, Langrod JG, eds. *Substance Abuse: A Comprehensive Textbook.* 2nd ed. Baltimore, Md: Williams & Wilkins; 1992:628–656.
27. Ostrea EM, Brady M, Gause S, Raymundo AL, Stevens M. Drug screening of newborns by meconium analysis: a large-scale, prospective, epidemiologic study. *Pediatrics.* 1992;89:107–113.
28. Phibbs CS, Bateman DA, Schwartz RM. The neonatal costs of maternal cocaine use. *JAMA.* 1991;266:1521–1526.
29. Lutiger B, Graham K, Einarson TR, Koren G. Relationship between gestational cocaine use and pregnancy outcome: a meta-analysis. *Teratology.* 1991;44:405–414.
30. Broekhuizen FF, Utrie J, Van Mullem C. Drug use or inadequate prenatal care? Adverse pregnancy outcomes in an urban setting. *Am J Obstet Gynecol.* 1992;166: 1747–1756.
31. Finnegan LP, Hagan T, Kaltenbach KA. Scientific foundation of clinical practice: opiate use in pregnant women. *Bull N Y Acad Med.* 1991;67:223–239.
32. Williams J. The real tragedy of crack babies. *Washington Post.* September 30, 1990:C5.

Black–White Differences in Serum Cotinine Levels among Pregnant Women and Subsequent Effects on Infant Birthweight

ABSTRACT

Objectives. Higher levels of serum cotinine (a metabolite of nicotine) have been found in Black smokers than in White smokers even after self-reported cigarette dose was controlled. It is unknown whether higher cotinine levels in Black pregnant smokers may increase the risk of delivering an infant of reduced birthweight.

Methods. We analyzed serum cotinine levels of 374 Black and 829 White women who smoked during pregnancy and who delivered between April 1964 and April 1967. Racial differences in the relationship between cotinine and birthweight were examined.

Results. Cotinine levels were 27.4 ng/mL higher in Black smokers after cigarette dose and confounding variables were controlled. Blacks had higher cotinine levels than Whites at each dose. No significant racial differences in the rate of decrease in birthweight per nanogram of cotinine per milliliter were found.

Conclusions. Our results confirm previous research showing higher cotinine levels at each smoking dose in Black smokers than in White smokers. Because there was no difference in the rate of decrease in birthweight due to cotinine, our results suggest that cigarette smoking among Blacks may have a greater effect on birthweight than it does among Whites. (*Am J Public Health.* 1994;84:1439–1443)

Paul B. English, PhD, MPH, Brenda Eskenazi, PhD, and Roberta E. Christianson, MA

Introduction

Black infants in the United States suffer from rates of low birthweight that are twice as high as those of White infants.[1] Although maternal smoking has been singled out as one of the most preventable risk factors for low birthweight,[2] little research has been conducted to examine whether Blacks are more susceptible to the reproductive hazards of cigarette smoke than Whites.[3,4]

Biochemical validation of smoking behavior in pregnant women by the measurement of cotinine, a major metabolite of nicotine, has been found to be a better predictor of birthweight than self-reported number of cigarettes per day.[5] In addition, serum cotinine levels in nonpregnant Black smokers have been found to be higher than those in White smokers even after controlling for a number of behavioral and personal characteristics, including self-reported cigarette dose.[6] Wagenknecht et al.[6] suggested that these higher cotinine levels in Blacks may reflect differences in nicotine absorption or metabolism. Others have suggested that these high cotinine levels may be due to racial differences in smoking characteristics or life-style.[7] It is unknown whether these racial differences persist in pregnant smokers and whether higher cotinine levels in Blacks, if observed, may result in an increased risk of delivering an infant of reduced birthweight.

We assessed the differences in serum cotinine levels among a cohort of Black and White pregnant women from the Oakland, Calif, area who, based on self-report, were smokers. The effects of these differences in cotinine levels on the birthweights of their infants were evaluated.

Methods

Participants

This study was based on a subsample of the Child Health and Development Studies, a longitudinal study that followed women and their offspring enrolled at the Kaiser Permanente Hospital facilities, primarily in Oakland, Calif. Pregnant women were enrolled in the study from 1959 to 1966. Child Health and Development Studies enrollees were socioeconomically diverse, with only the extremes—the very impoverished or the very affluent—not being well represented.[8]

Women were questioned about their smoking status during their first Child Health and Development Studies interview, which usually took place in the first trimester at a prenatal care visit. Serum samples were taken from the gravida in each of the three trimesters and at postpartum. A second interview was conducted in the second half of pregnancy to assess whether any differences in health, attitudes, and behavior had occurred. At the first interview (and second interview if data were available), women who reported that they smoked were classified as smokers; women who had never smoked or who had quit before pregnancy were

The authors are with the Maternal and Child Health and Epidemiology Programs, School of Public Health, University of California, Berkeley, Calif.

Requests for reprints should be sent to Paul B. English, PhD, MPH, 312 Warren Hall, School of Public Health, University of California, Berkeley, Berkeley, CA 94720.

This paper was accepted November 15, 1993.

TABLE 1—Sociodemographic Characteristics and Pregnancy History of Black and White Pregnant Smokers and Nonsmokers, Child Health and Development Studies, Oakland, Calif, 1964 through 1967

Characteristic	Black		White	
	Nonsmokers (n = 696)	Smokers (n = 374)	Nonsmokers (n = 1444)	Smokers (n = 829)
Age, y (mean ± SD)	27.3 ± 6.3	26.7 ± 6.3	27.1 ± 5.6	26.3 ± 5.4
Body mass index, %				
Low	12.6	20.3	19.6	24.4
Medium	57.2	57.5	68.9	63.3
Heavy	13.8	10.7	4.9	3.7
Obese	11.9	8.8	3.5	5.4
Unknown	4.5	2.7	3.0	3.1
Parity, %				
Primiparous	24.1	26.2	41.0	35.0
Low multiparous (1–2)	44.7	41.2	47.2	47.5
High multiparous (≥3)	31.2	32.6	11.8	17.5
Education, %				
Did not graduate from high school	18.1	30.7	8.4	21.2
High school graduate	76.1	64.4	61.2	61.4
College graduate	5.6	4.8	30.3	17.4
Unknown	0.1	...	0.1	...

as nonsmokers. One serum sample was chosen for each woman. The sample used was the closest by date to the average time of the second interview because this time was considered the most critical period for effects on fetal growth.[2] (The mean time for the second interview was 185 days after the woman's last menstrual period.)

Smoking consumption was recorded in the interviews in categories of 1–4, 5–9, 10–14, 15–19, 20–29, 30–39, 40–60, and more than 60 cigarettes per day. For some analyses, level of smoking consumption was recategorized as 1–9, 10–19, and 20 or more cigarettes per day from the initial and the second Child Health and Development Studies interviews. Other smoking information (e.g., the number of years the mother smoked, whether she inhaled and/or smoked nonfilter cigarettes) was also obtained from the initial interview. Other first-interview variables included maternal educational achievement (defined as did not graduate from high school, high school graduate, and college graduate), parity (defined as the total number of previous live births and stillbirths), age, prepregnant weight, height, and race. Body mass index was categorized into four levels: low (<19.8 kg/m^2), medium (19.8–26.0 kg/m^2), heavy (26.1–29.0 kg/m^2), and obese (>29.0 kg/m^2). Data from the second interview were used for the category of number of cigarettes smoked, if this information was available; otherwise, it was taken from the initial interview. Information on birthweight (in ounces) of the infant was obtained from the infant's Kaiser medical chart; gestational age was estimated from last menstrual period as reported in the interview.

Our original sample consisted of 4240 pregnancies (live births and fetal deaths) of Black and White women who were interviewed during pregnancy and delivered between April 1964 and April 1967. We excluded plural births (n = 64), pregnancies with unknown smoking status of the mother (n = 20), and pregnancies with unavailable sera for testing (n = 27). We further excluded those with unknown birthweight or gestational age (n = 142) and those where the mother either started or stopped smoking during pregnancy (n = 397). An additional 247 pregnancies were dropped to limit the sample to one pregnancy per woman; the sibling retained was chosen randomly. The final sample available for analysis was 3343 women, with 374 smokers and 696 nonsmokers among Blacks and 829 smokers and 1444 nonsmokers among Whites. Of these 3343, 68.8% had completed a second interview.

Cotinine Analysis

Once nicotine is absorbed into human tissue via smoking, it is converted to cotinine and other metabolites in a series of steps of oxidative metabolism.[9] Nicotine acts as a vasoconstrictor in both the mother and fetus and is thought to be one of the main compounds in tobacco smoke to induce intrauterine growth retardation.[9] About 70% of nicotine is converted to cotinine, which can be detected in serum.[10] Cotinine has a half-life of 15–20 hours in adults.[11] Frozen 3-cc samples of serum were stored either at a cold storage facility in Oakland, Calif (22.4% of all samples) or in Frederick, Md (77.6% of all samples). The samples were shipped from both facilities to the Clinical Biochemical Laboratories at the American Health Foundation in Valhalla, NY, for analysis. The samples used were collected between 1964 and 1967.

Cotinine was determined by the radioimmunoassay method of Haley et al.[12] after a modification of a technique by Langone et al.[13] The interassay coefficient of variation for cotinine was found to be 10.9%, with a minimal detection limit of 2 ng/mL. Out of the entire population of samples, 10% were blindly selected for duplicate analysis. Cross-tabulation of the duplicate samples of smokers and nonsmokers over five categories[6] of cotinine levels (0–13, 14–99, 100–249, 250–499, and ≥500 ng/mL) yielded an agreement rate of 97%.

Data Analysis

Comparison of median cotinine values by race was done by using the Wilcoxon rank sum test, and confidence intervals (CIs) for the medians were computed based on the binomial method.[14] Spearman correlation coefficients were calculated between the ranked categories of self-reported cigarettes (using the eight-category classification) and serum cotinine. Mantel-Haenszel chi-square values were calculated for categorical data, and comparison of means was done by using *t* tests. Gestational age later than 40 weeks was classified as 40 weeks for analysis.[15]

Simple linear regression models were used for each race and at each dose level to examine the effect of cotinine on birthweight. Because models using the log transformation of cotinine and categorizing cotinine in 100-ng/mL intervals did not explain as much of the variance in birthweight as untransformed cotinine, untransformed cotinine was used in the analysis. Multivariate linear regression models were first run separately for each race for all smokers and at each dose level. Additional multivariate models were run to test for interaction by fitting models with a cross-product term for race and cotinine ($P < .10$). SAS and Confidence

Interval Analysis software was used in the analyses.[16,17]

Results

Table 1 shows the characteristics of women in the study sample stratified by race. Average age was similar for White and Black women ($t = -1.33$, $P = .18$), whereas more White women than Black women were college graduates ($\chi^2 = 210.0$, $P < .001$) and primiparae ($\chi^2 = 160.9$, $P < .001$). More Blacks than Whites had a heavy or obese body mass index ($\chi^2 = 143.7$, $P < .001$). Among both races, more smokers than nonsmokers did not graduate from high school ($\chi^2 = 22.2$, $P < .001$ for Blacks; $\chi^2 = 101.1$, $P < .001$ for Whites).

Analysis of smoking behavior (Table 2) showed that Blacks were more likely than Whites to report smoking fewer cigarettes per day. Blacks were also less likely to inhale and to use nonfilter cigarettes. The number of years smoked and the percentage of women who had husbands who smoked were similar in both races.

Cotinine Levels

Out of all participants who classified themselves as smokers at the interviews (n = 1203), 93.3% (n = 1123) had serum cotinine levels greater than or equal to 14 ng/mL (cutoff level for active smoking[18]). Out of all participants who reported that they were nonsmokers at the interviews (n = 2140), only 2.0% (n = 42) had cotinine levels greater than or equal to 14 ng/mL (false-negative rate). False-negative and positive predictive value rates were 2.9% and 95.5% for Blacks and 1.5% and 92.4% for Whites.

Because we were concerned about sample desiccation given the age of the samples, we examined sodium levels in a random sample of 50 serums from each facility. Average sodium levels were slightly higher than normal (normal range = 137–145 mmol/L) in samples from the Frederick facility (mean = 153.72 mmol/L, SD = 21.59) and even higher in samples from the Oakland facility (mean = 189.38 mmol/L, SD = 36.46). To test the impact of including Oakland samples in this study, we computed the difference in average cotinine levels among smokers, including and excluding samples from the Oakland facility, for each racial group. When excluding these samples, average cotinine levels were only 3% (136.5–132.4 ng/mL) and 1.7% (147.6–145.1 ng/mL) lower in White and Black smokers, respectively (proportions of Whites and Blacks were similar in each facility). These results indicate that there is little evidence of systematic bias of sample desiccation on our analysis by race.

Cotinine levels agreed well with self-reported smoking status in that the median cotinine levels increased with the reported category of the number of cigarettes for Blacks and for Whites (Table 2). The Spearman correlation coefficients for ranked categories of cigarettes and cotinine were 0.52 ($P = .0001$) for Whites and 0.37 ($P = .001$) for Blacks. Previous studies have found correlations between self-reported cigarettes and cotinine of 0.58 (both races),[19] 0.64 in Whites and 0.46 in Blacks.[6]

Although Black smokers reported smoking fewer cigarettes per day, they had higher mean ($P = .05$) and median ($P = .13$) serum cotinine levels than White smokers (Table 2). Black smokers also had higher cotinine levels in each category of self-reported cigarettes smoked per day. We also found that 18.4% of Black smokers compared with 3.9% of White smokers exceeded levels of 25 ng of cotinine/mL per cigarette (cotinine levels per cigarette were derived by using the midpoint of each category of cigarette consumption).[20]

After number of years smoked, number of cigarettes smoked, whether or not the participants used nonfiltered cigarettes, education, and age were controlled for in a multivariate regression model, cotinine levels were 27.4 ng/mL higher in Blacks than in Whites. Cotinine levels stayed significantly higher in Blacks after

TABLE 2—Serum Cotinine Levels and Smoking Characteristics of Black and White Pregnant Smokers, Child Health and Development Studies, Oakland, Calif, 1964 through 1967

Characteristic	Blacks (n = 374)	Whites (n = 829)	χ^2, t, or z	P
Self-reported cigarettes/d, %				
1–9	40.9	22.7	$\chi^2 = 62.5$	<.001
10–19	32.9	28.7		
≥20	26.2	48.5		
Unknown	...	0.1		
Inhales, %	92.5	95.5	$\chi^2 = 4.61$	.03
Smokes nonfilter cigarettes, %	11.2	15.6	$\chi^2 = 3.96$	.05
Years smoked, mean ± SD	8.67 ± 5.53	8.88 ± 5.0	$t = 0.65$	.52
Husband smokes, %	69.8	66.5	$\chi^2 = 1.42$	.49
Median serum cotinine, ng/ML (95% CI)	135(119,147)	130(121,137)	$z = 1.51$	.13
Serum cotinine, mean ± SD	147.6 ± 97.3	136.5 ± 90.5	$t = 1.92$	.05
Median serum cotinine for self-reported cigarettes/d, ng/mL (95% CI)				
1–9	91(73,110)	44(33,54)	$z = 5.95$	.0001
10–19	149(134,176)	128(112,141)	$z = 2.72$	.006
≥20	182(150,207)	164(152,174)	$z = 1.19$	.23

TABLE 3—Birthweight and Gestational Age of Infants of Black and White Smokers and Nonsmokers, by Self-Reported Cigarettes per Day

Self-Reported Smoking Status	Mean ± SD	
	Birthweight, g	Gestational Age, d
Black (n = 1070)		
Nonsmokers	3241.2 ± 573.7	277.7 ± 20.9
1–9 cigarettes/d	3030.3 ± 636.3	272.6 ± 26.5
10–19 cigarettes/d	3026.7 ± 594.9	276.2 ± 21.2
≥20 cigarettes/d	2964.0 ± 637.2	273.7 ± 25.7
White (n = 2273)		
Nonsmokers	3426.2 ± 565.0	282.4 ± 17.2
1–9 cigarettes/d	3295.3 ± 579.5	283.0 ± 18.4
10–19 cigarettes/d	3274.6 ± 531.5	281.4 ± 18.2
≥20 cigarettes/d	3218.9 ± 482.8	281.8 ± 16.5

TABLE 4—Results of Multivariate Linear Regression[a] of the Effect of Maternal Serum Cotinine on Infant Birthweight, by Race, Child Health and Development Studies, Oakland, Calif, 1964 through 1967

	Mean Difference in Birthweight (g) per ng of Cotinine/mL (95% CI)		
	Black (n = 364)	White (n = 806)	P[b]
All smokers	−1.18 (−1.69, −0.67)	−0.85 (−1.21, −0.50)	.10
Cigarettes/d			
1–9	−1.50 (−2.49, −0.52)	−1.22 (−2.34, −0.11)	.80
10–19	−1.08 (−2.01, −0.15)	−1.73 (−2.55, −0.91)	.38
≥20	−0.82 (−1.79, 0.15)	−0.41 (−0.92, 0.10)	.12

[a]Variables controlled for were education (did not graduate from high school, high school graduate vs college graduate [reference]), maternal age, parity (primiparous, high multiparous vs low multiparous [reference]), gestational age, inhalation, filter cigarette use, number of years smoked, and body mass index (medium, heavy, obese vs low [reference]). Ten records for Blacks and 23 records for Whites were deleted because of missing values.
[b]P value for cross-product term for interaction between race and serum cotinine.

conducting this analysis separately by educational level (not shown).

Relationship of Cotinine to Birthweight

Black women delivered infants weighing on the average 202 g less than infants delivered by White women. Table 3 shows the effect in both races of self-reported smoking on birthweight. Black women smoking 20 or more cigarettes per day delivered infants that weighed on average 277 g (95% CI = 154, 400) less than those delivered by nonsmokers; among Whites, the deficit was 207 g (95% CI = 147, 268).

Among all smokers, the results of simple linear regression (not shown) revealed that cotinine levels were associated with a decrease of 1.13 g per ng of cotinine/mL (95% CI = −1.47, −0.79). Serum cotinine was somewhat more related to birthweight among Blacks than among Whites (−1.22 g per ng/mL [95% CI = −1.86, −0.58] vs −0.98 g per ng/mL [95% CI = −1.37, −0.60]). After stratification by self-reported cigarettes per day, the effect of cotinine on birthweight was found to be larger in White light (1–9 cigarettes/day) smokers than in Black light smokers (−2.30 g per ng/mL vs −1.44 g per ng/mL). However, Black heavy (≥20 cigarettes/day) smokers showed a larger effect of cotinine on birthweight than White heavy smokers (−1.38 g per ng/mL vs −0.48 g per ng/mL).

To examine these racial differences further, we assessed the effect of the amount of cotinine on birthweight while controlling for parity, education, maternal age, body mass index, and smoking behavior variables such as inhalation, nonfilter cigarette use, and the number of years smoked. Table 4 shows the relationship between infant birthweight and serum cotinine derived from multivariate linear regression models run for each race. Among all smoking dose levels, an increase of 1 ng of cotinine/mL resulted in a decrease of 0.85 g (95% CI = −1.21, −0.50) among Whites and a decrease of 1.18 g (95% CI = −1.69, −0.67) among Blacks (*P* value for interaction between race and cotinine = .10). After conducting separate interaction models by self-reported cigarettes per day, no evidence for an interaction (at $P < .10$) between race and cotinine level was found (Table 4).

Discussion

In a study of young Black and White smokers, Wagenknecht et al. found that Blacks had cotinine levels that were 83.3 ng/mL higher than those of Whites after controlling for several characteristics, including number of cigarettes, inhalation frequency, age, and education.[6] After controlling for similar factors in our study of pregnant women, we found that Black smokers had cotinine levels that were 27.4 ng/mL higher than those of White smokers. Like Wagenknecht et al., we also found that Black smokers had higher serum cotinine levels at each level of self-reported cigarette consumption.

Although not statistically significant in this study, the effect on birthweight of self-reported smoking appeared to be greater in Blacks than in Whites. Blacks smoking a pack or more of cigarettes a day delivered infants that weighed on average 277 g less than those delivered by nonsmokers. The birthweight decrement in Whites smoking a similar amount was only 207 g.

After analyzing the relationship between serum cotinine and birthweight, we found little evidence for a racial difference in the *rate* of reduction in birthweight per ng of cotinine/mL, even after controlling for maternal factors and smoking behavior including self-reported cigarette consumption (Table 4). However, Blacks had higher levels of cotinine at each category of cigarette dose (Table 2). Thus, these data suggest that the per-cigarette effect on birthweight may be greater in Blacks than in Whites.

The higher average cotinine levels among Blacks may be due to several factors. First, it has been previously suggested that the higher levels may be due to differences in nicotine metabolism.[6] The amount of cotinine measured in serum is dependent not only on the original dose of nicotine absorbed and the rate of conversion of nicotine to cotinine, but also on the rate of excretion.[10] Factors such as exercise and physical fitness may also affect nicotine metabolism.[7] Although we were able to control for body mass index as a measure of fitness, we were unable to assess whether racial differences in exercise or cotinine excretion rates exist in this study.

Second, a possible explanation for this racial difference could be differential self-report. In a controlled study,[21] heavy smokers that were restricted to only a few cigarettes (and therefore expected to have a higher intake of smoke per cigarette) did not exceed 25 ng of cotinine/mL per cigarette. It has been suggested, therefore, that smokers that exceed this level may have underreported their cigarette consumption.[20] We found that 4.7 times as many Black smokers as White smokers exceeded levels of 25 ng of cotinine/mL per cigarette. In addition, we found that the false-negative rate (percentage of nonsmokers with cotinine levels of ≥14 ng/mL) in Blacks was almost double the rate of Whites (2.9% vs 1.5%, respectively).

In contrast, we found that Black smokers continued to have higher cotinine levels than White smokers after stratifying by educational group. Any underreporting would be expected to vary by educational level.[6] Further, no evidence of a systematic racial bias in self-report of cigarette consumption has

been reported in the National Health Interview Survey,[22] and Wagenknecht et al. have estimated that Black smokers in their study would have to smoke twice as much as reported to explain the race differential in cotinine levels.[6]

Finally, differences in cotinine levels may also be due to racial variations in smoking characteristics. In this study, more Whites than Blacks were heavy smokers, inhaled, and smoked nonfiltered cigarettes. Therefore, we would have expected that White smokers would have greater exposure to nicotine than Black smokers and hence higher, not lower, cotinine levels. However, we lacked information on smoking behavior such as brand preference, mentholated cigarette consumption, and length of cigarette smoked, which may affect nicotine exposure.[7] It has been reported that Black smokers actually smoke less of their cigarettes than White smokers.[23] Further, Black smokers have been previously found to have higher serum cotinine levels than White smokers even after controlling for mentholation of the cigarette.[6]

In conclusion, Black pregnant smokers in this study were found to have higher serum cotinine levels than White smokers after controlling for smoking dose and demographic confounders. Given these higher levels and the fact that we found no racial difference in the rate of decrease in birthweight per nanogram of cotinine per milliliter, these data suggest that cigarette smoking among Blacks may have a greater effect on birthweight than it does among Whites. Experimental studies are needed to better characterize racial differences in nicotine exposure and cotinine levels. □

Acknowledgments

This study was supported by funds provided by the Cigarette and Tobacco Surtax Fund of the state of California through the Tobacco-Related Disease Research Program of the University of California (grant number 1RT465).

We thank the Child Health and Development Studies for the use of their data. Drs Neil Benowitz, Robert Engel, Jeffrey Gould, Nancy Haley, Martin Kharrazi, William Satariano, and Lynne Wagenknecht made helpful suggestions on an earlier draft. Caryn Axelrad of the Clinical Biochemistry Facility of the American Health Foundation conducted the cotinine analysis.

References

1. *Child Health USA.* Washington, DC: US Dept of Health and Human Services, Health Resources and Services Administration; 1989.
2. *Preventing Low Birthweight.* Washington, DC: Institute of Medicine, National Academy Press; 1985.
3. Alameda County Low Birth Weight Study Group. Cigarette smoking and the risk of low birth weight: a comparison in black and white women. *Epidemiology.* 1990;1:201–205.
4. Kleinman JC, Pierre MB Jr, Madans JH, et al. The effects of maternal smoking on fetal and infant mortality. *Am J Epidemiol.* 1988;127:274–282.
5. Haddow JE, Knight GJ, Palomaki GE, et al. Cigarette consumption and serum cotinine in relation to birthweight. *Br J Obstet Gynaecol.* 1987;94:678–681.
6. Wagenknecht LE, Cutter GR, Haley NJ, et al. Racial differences in serum cotinine levels among smokers in the coronary artery risk development in adults study. *Am J Public Health.* 1990;80:1053–1056.
7. McCarthy WJ, Caskey NH, Jarvik ME. Ethnic differences in nicotine exposure. *Am J Public Health.* 1992;82:1171–1172. Letter.
8. van den Berg B. The California Child Health and Development Studies. In: Mednick S, ed. *Handbook of Longitudinal Research.* New York, NY: Praeger Publishers; 1984:166–179.
9. Nash JE, Persaud TV. Embryopathic risks of cigarette smoking. *Exp Pathol.* 1988;33: 65–73.
10. Jacob P, Benowitz NL, Shulgin AT. Recent studies of nicotine metabolism in humans. *Pharmacol Biochem Behav.* 1988;30:249–253.
11. Benowitz NL, Kuyt F, Jacob P, et al. Cotinine disposition and effects. *Clin Pharmacol Ther.* 1983;34:604–611.
12. Haley N, Axelrad C, Tilton K. Validation of self-reported smoking behavior: biochemical analyses of cotinine and thiocyanate. *Am J Public Health.* 1983;73:1204–1207.
13. Langone JJ, Gjida HB, Van Vunakis H. Nicotine and its metabolites. Radioimmunoassays for nicotine and cotinine. *Biochemistry.* 1973;12:5025–5030.
14. Gardner MJ, Altman DG, eds. *Statistics with Confidence.* London, England: British Medical Journal; 1989.
15. Showstack JA, Budetti PP, Minkler D. Factors associated with birthweight: an exploration of the roles of prenatal care and length of gestation. *Am J Public Health.* 1984;74:1003–8.
16. *SAS/STAT User's Guide, Release 6.03 Edition.* Cary, NC: SAS Institute Inc; 1988.
17. Gardner SB, Winter PD, Gardner MJ. *Confidence Interval Analysis, Version 1.0.* London, England: British Medical Journal; 1989.
18. Cummings SR, Richard RJ. Optimum cutoff points for biochemical validation of smoking status. *Am J Public Health.* 1988;78: 574–575.
19. Pattishall EN, Strope GL, Etzel RA, et al. Serum cotinine as a measure of tobacco smoke exposure in children. *Am J Dis Child.* 1985;139:1101–1104.
20. Perez-Stable EJ, Marin BV, Marin G, et al. Apparent underreporting of cigarette consumption among Mexican American smokers. *Am J Public Health.* 1990;80:1057–1061.
21. Benowitz NL, Jacob P, Kozlowski LT, et al. Influence of smoking fewer cigarettes on exposure to tar, nicotine, and carbon monoxide. *N Engl J Med.* 1986;315:1310–1313.
22. Sterling TD, Weinkam JJ. Comparison of smoking-related risk factors among Black and White males. *Am J Ind Med.* 1989;15: 319–333.
23. Wagenknecht LE, Haley NJ, Jacobs DR. Wagenknecht and colleagues respond. *Am J Public Health.* 1992;82:1173. Letter.

Birthweight Differentials among Asian Americans

Gopal K. Singh, PhD, MS, and Stella M. Yu, ScD, MPH

ABSTRACT

Objectives. This study examines differentials in mean birthweight and the risk for low birthweight among various Asian-American groups in New York State (n = 499 377).

Methods. Using resident singleton live-birth records from New York State for 1985 and 1986, Asian-American births were compared with Black, American Indian, and White births. Multivariate ordinary least squares and logistic regression models were used to analyze ethnic differences.

Results. Compared with White births, the expected mean difference in birthweight was −115 g for Chinese, −235 g for Japanese, −164 g for Filipinos, −120 g for Blacks, and 74 g for American Indians. The risk for low birthweight was 45% higher for Filipinos and 49% higher for Blacks as compared with Whites.

Conclusions. Results of this study suggest substantial heterogeneity in mean birthweight and risk for low birthweight among ethnic groups in general and the major Asian-American groups in particular. Interestingly, after controlling for ethnic differences in sociodemographic risk factors, Filipinos appear to resemble Blacks much more closely than they do their Japanese and Chinese counterparts with respect to risk for low birthweight. (*Am J Public Health.* 1994;84:1444–1449)

Introduction

The significance of birthweight as one of the most important determinants of infant mortality and morbidity is well documented in social and biomedical research.[1] For infants of low birthweight (i.e., less than 2500 g), the risk of dying in the first year of life is about 21 times that for infants weighing 2500 g or more. For infants of very low birthweight (i.e., less than 1500 g), the risk is considerably greater: about 90 times that for infants weighing 2500 g or more.[2] According to the most recent vital statistics data, the rate of low birthweight (percentage of newborn babies weighing less than 2500 g) in the United States was 7.1% in 1991. However, this rate varied substantially among the various racial and ethnic groups. For White infants it was 5.8%, for Black infants it was 13.6%; and for American Indians and Asian-American groups such as Japanese, Filipinos, and Hawaiians, it was 6.2%, 5.9%, 7.3%, and 6.7%, respectively, all somewhat higher than that observed for White infants. The rate was lowest for Chinese infants (5.1%).[3]

Given these statistics and the particularly strong impact of birthweight on infant mortality and morbidity, it is not surprising that numerous studies have attempted to analyze risk factors associated with low birthweight. However, there is a paucity of studies that have examined racial/ethnic differentials in birthweight among US populations other than Whites, Blacks, and Hispanics.[4] Especially lacking are birth outcome studies on specific Asian-American subgroups, some of which are among the fastest growing minority groups in the United States.[5,6]

Despite the presence of marked heterogeneity among Asian Americans, studies investigating ethnic differentials in socioeconomic characteristics and birth outcomes have often treated Asian Americans as a single entity.[7–9] However, previous research using linked birth and infant death records has demonstrated important variations among Asian Americans in infant mortality rates by age at death[10–15] and cause of death.[16] Similarly, profound differences among these people have been noted with respect to low birthweight and associated risk factors.[4,11] Results of this paper further illustrate the substantial heterogeneity that exists within the Asian-American population with respect to sociodemographic characteristics and birth outcomes.

The purpose of this study is to examine ethnic differentials in mean birthweight and risk for low birthweight, with special emphasis on three major Asian-American groups: Chinese, Japanese, and Filipino Americans. We evaluate birth outcomes among these three groups relative to those for White, Black, and American Indian infants. The specific questions we investigate are (1) whether these racial/ethnic groups differ significantly with respect to mean birthweight and risk of low birthweight, and (2) whether these differentials, if they exist, persist after controlling for selected sociodemographic and biological risk factors.

Gopal K. Singh is with the National Center for Health Statistics, Division of Vital Statistics, Centers for Disease Control and Prevention, Hyattsville, Md. Stella M. Yu is with the Maternal and Child Health Bureau, Health Resources and Services Administration, Rockville, Md.

Requests for reprints should be sent to Gopal K. Singh, PhD, MS, Centers for Disease Control and Prevention, National Center for Health Statistics, Division of Vital Statistics, 6525 Belcrest Rd, Room 840, Hyattsville, MD 20782.

This paper was accepted January 10, 1994.

Data and Methods

Data

The data used in this study are derived from the National Linked Birth/Infant Death Data Sets for 1985 and 1986 birth cohorts.[2,17] From these two national files were extracted all the resident singleton live births that occurred in New York State during those 2 years.

For the purpose of this analysis, White, Black, American Indian, Chinese, Japanese, and Filipino live births in New York State in 1985 and 1986 were considered. The births occurring to other ethnic groups, including other Asian Americans, were excluded because of insufficient numbers of live births for analysis. The study sample for New York State thus consisted of 499 397 singleton live births, with Chinese, Japanese, and Filipino births in the sample accounting for about 1.5% of all births in the state. It is important to note that all live births were classified according to maternal race/ethnicity.

The rationale for selecting New York State was threefold. First, it is impractical and costly to conduct a multivariate analysis at the national level using all 3.7 to 4.1 million births that occur during a specific calendar year in the United States. Second, although a large proportion (about 40%) of Asian-American births occur in California, that state could not be included in the analysis because information on maternal education—an important covariate of birth outcome and the only measure of socioeconomic status available from vital statistics data—is not recorded on its birth certificates. Third, the population of New York State is ethnically quite diverse, yielding sufficiently large numbers of live births for each of the ethnic groups considered here.

TABLE 1—Selected Maternal and Infant Characteristics Used in the Analysis of Differentials in Mean Birthweight and Risk for Low Birthweight, by Race/Ethnicity: New York State, 1985 to 1986

	Race/Ethnicity					
	White	Black	American Indian	Chinese	Japanese	Filipino
Birthweight						
Mean, g	3396.1	3143.3	3413.7	3283.9	3221.5	3244.3
SD	557.8	624.4	610.9	454.9	456.6	487.8
% low birthweight	5.0	11.5	5.4	3.9	4.0	5.7
Maternal age, y						
<20	7.9	17.4	20.5	0.9	1.0	1.1
20–34	82.9	74.5	74.9	85.8	78.6	77.1
35+	9.2	8.1	4.6	13.3	20.4	21.8
Maternal education, y						
<12	16.5	32.1	36.7	18.1	3.7	4.7
12	41.1	41.6	40.0	51.6	23.5	18.3
13+	42.4	26.3	23.3	30.3	72.8	77.0
Mean	12.9	11.9	11.6	12.4	14.4	14.9
Marital status						
Unmarried	19.5	63.6	49.9	4.4	4.0	10.3
Married	80.5	36.4	50.1	95.6	96.0	89.7
Nativity status						
Native born	81.6	66.4	84.0	5.6	7.0	2.5
Foreign born	18.4	33.6	16.0	94.4	93.0	97.5
County of residence						
Metropolitan	73.1	97.2	60.1	97.4	95.9	96.8
Nonmetropolitan	26.9	2.8	39.9	2.6	4.1	3.2
Total birth order						
1	34.7	33.3	30.0	47.7	39.5	51.6
2–3	49.3	43.7	45.2	44.5	54.4	40.2
4+	16.0	23.0	24.8	7.8	6.1	8.2
Prenatal care began						
1st trimester	78.0	51.9	63.7	65.0	84.7	71.8
2nd–3rd trimester	18.9	37.0	34.8	32.5	14.1	24.2
No care	3.1	11.1	1.5	2.5	1.2	4.0
Mean no. prenatal visits	10.4	7.7	9.5	8.7	11.0	9.1
Gestational age, wk						
<37	9.1	15.8	12.8	6.5	6.6	8.8
37+	90.9	84.2	87.2	93.5	93.4	91.2
Live births, n	382 506	10 830	927	5757	531	1356

Source: National Center for Health Statistics, *Linked Birth/Infant Death Data Sets*, 1985 and 1986 birth cohorts.[2,17]

Variables

The dependent variables in this study were mean birthweight and risk of low birthweight. In estimating the risk of low birthweight, all the live births weighing less than 2500 g were defined to be of low birthweight.[18] Differences in these two dependent variables were modeled as functions of both sociodemographic and biological risk factors. More specifically, the risk factors (covariates) considered to influence birthweight included maternal background factors (e.g., age, race/ethnicity, educational attainment, nativity, marital status, county of residence, and prenatal care) and infant characteristics (e.g., birth order, sex, and gestational age).[19] Maternal education and prenatal care were measured as continuous variables, while all other covariates, including race/ethnicity, were measured as categorical variables.

Methods

Ordinary least squares regression was used to analyze ethnic differences in mean birthweight while multiple logistic regression was used to analyze ethnic differences in the risk of low birthweight, controlling for the effects of other covariates. The parameters in the ordinary least squares model, which represented the effects of covariates including ethnicity, were estimated using the REG procedure; those in the logistic model were estimated by the maximum likelihood method using the LOGISTIC procedure of SAS, Version 6.[20]

Racial/ethnic differentials in mean birthweight and risk of low birthweight were examined by fitting a series of hierarchical regression models, keeping in mind the causal sequencing of the risk factors in their associations with birthweight. In the first stage, regression models were fitted to determine crude (unadjusted) differentials in mean birth-

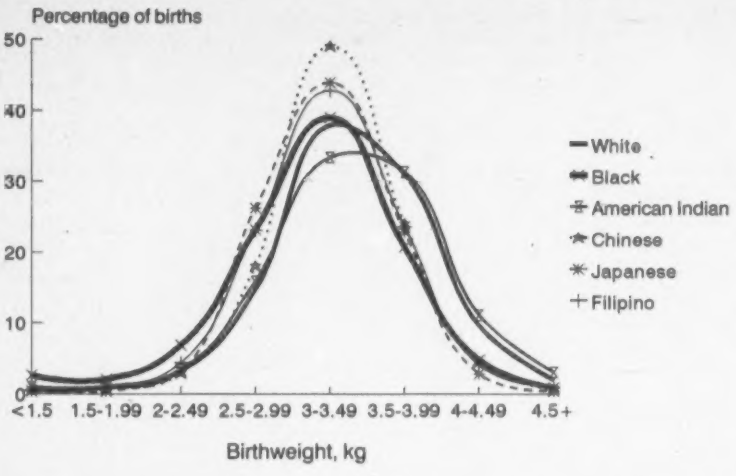


FIGURE 1—Birthweight distribution by maternal race/ethnicity: New York State, 1985 to 1986.

weight and risk of low birthweight between Whites and each of the other ethnic groups. In the second stage, ethnic differentials were investigated by entering race/ethnicity, maternal age, education, marital status, nativity, place of residence, birth order, infant sex, and prenatal care into the model. The final model examined ethnic differentials, incorporating all the covariates including gestational age.

Results

Table 1 shows descriptive data on birth outcomes and selected maternal and infant characteristics by race/ethnicity. As seen in this table, Chinese, Japanese, and Filipino infants weighed significantly less at birth (about 120 to 180 g less) than White infants. Black infants had the lowest mean birthweight, approximately 250 g less than their White counterparts. American Indians had on average about the same birthweight as Whites. However, Chinese, Japanese, and Filipino Americans had a much more compact birthweight distribution, showing substantial clustering around the mean birthweight, whereas White, Black, and American Indian babies showed considerable dispersion below and above the mean (see Figure 1). This also implies that variation in birthweight among Asian Americans was substantially lower than that among Whites, Blacks, and American Indians, which is also evident by looking at the differences in standard deviations of race/ethnic-specific birthweight shown in Table 1.

According to the figures in Table 1, Chinese had the lowest rate of low birthweight, about 22% lower than the White rate and 32% lower than the Filipino rate. Rates of low birthweight for Japanese, Filipinos, and American Indians were not significantly different from that for Whites. However, Blacks had the highest rate—about 2.3 times that for Whites and nearly three times that for Chinese.

The groups with the highest rates of low birthweight appear to have the least favorable distributions on several of the risk factors (see Table 1). For example, the percentage of births to mothers aged 19 or less, the incidence of nonmarital childbearing (as measured by the proportion of births occurring to unmarried mothers), and the proportion of fourth- and higher-order pregnancies were substantially higher among Blacks, American Indians, and Whites than among Asian groups. Similarly, the proportion of preterm births was higher among Blacks, American Indians, and Whites than among Chinese, Japanese, and Filipinos. Additionally, not only did Black and American Indian mothers have, on average, fewer years of schooling than their Asian counterparts, but they were also significantly less likely to have attained a college education, and a somewhat lower proportion of them sought prenatal care during the first trimester of pregnancy.

Ethnic Differentials in Mean Birthweight

Results of multiple linear regressions examining the effects of race/ethnicity and other covariates on mean birthweight are presented in Table 2. When only race/ethnicity was considered, all groups except American Indians differed significantly from Whites in mean birthweight (model 1). When maternal age, education, marital status, nativity, county of residence, birth order, infant sex, and prenatal visits were introduced as controls in model 2, the Black/White differential in mean birthweight decreased by 108 g and the Chinese/White differential decreased by 8 g, while the Japanese/White, Filipino/White, and American Indian/White differentials increased by 67, 24, and 55 g, respectively. In other words, the standardized differences in mean birthweight between Whites and Japanese, Filipinos, and American Indians were larger and those between Whites and Blacks were significantly smaller than the corresponding crude observed differences in model 1.

The other covariates in model 2 were related to birthweight in the manner expected. Maternal age of 35 years and older; first-order, female, and out-of-wedlock births; mother's metropolitan residence, US-born status, and lower educational attainment; and fewer prenatal visits were all associated with significantly lower mean birthweights. Model 3 shows the ethnic differentials in mean birthweight after controlling for the additional effect of gestational age. Relative mean differences in birthweight among ethnic groups changed only slightly, revealing essentially similar patterns as those observed in model 2. Compared with White births, the expected mean difference in birthweight was −115 g for Chinese, −235 g for Japanese, −164 g for Filipinos, −120 g for Blacks, and 74 g for American Indians.

Ethnic Differentials in Risks for Low Birthweight

Since the analysis of ethnic differentials in mean birthweight may be affected by ethnic differences at the lower and upper extremes of the birthweight distribution, a separate analysis of low birthweight is warranted to identify risk factors associated with ethnic differentials in birth outcome. The results of this analysis are presented in Table 3. Clearly, the results of Table 2 are reinforced by the ethnic differentials in the risk for low

birthweight estimated by logistic regression in models 1 through 3 of Table 3. Model 1 in Table 3 presents the crude effect of race/ethnicity on the risk for low birthweight. The category showing an odds ratio of 1.00 represents the reference group used for comparison. Compared with White infants, Black infants had a 148% higher risk for low birthweight whereas Chinese had a 22% lower risk. However, the differences in low-birthweight rates between Whites and Japanese, Filipinos, and American Indians were not statistically significant. Not surprisingly, the results in model 1 of Table 3 are consistent with those reported earlier while discussing the ethnic differentials in low birthweight rates in Table 1.

To further explain ethnic differentials in risks for low birthweight, we added to model 1 sociodemographic risk factors, as mentioned above. Results of this stage are shown in model 2. Controlling for the confounding effects of these covariates reduced ethnic difference in risks for low birthweight for Black infants from 148% to 59% but increased significantly the difference for Filipino infants from 15% to 54%. However, differences in risks for low birthweight between Whites and Japanese, Chinese, and American Indians were not statistically significant.

The remaining results in model 2 of Table 3 are consistent with expectations concerning the relationships of the covariates considered and the risks for low birthweight. Births to unmarried mothers as well as to mothers aged 35 and over, out-of-wedlock births, first- and fourth- or higher-order pregnancies, and births to US-born mothers were all associated with a significantly higher risk for low birthweight. Furthermore, that risk declined significantly with increasing levels of education and prenatal care, even after controlling for the effects of other covariates.

Next added to model 2 was gestational age, the findings of which are shown in model 3 of Table 3. Compared with the other risk factors, gestational age had the strongest net impact on the risk for low birthweight. Although its inclusion in the model somewhat narrowed the ethnic differentials in low birthweight, Filipino and Black infants continued to show significantly higher risks for low birthweight than Whites—45% and 49%, respectively.

TABLE 2—Ordinary Least Squares Regressions Showing Crude and Adjusted Racial/Ethnic Differentials in Mean Birthweight: New York State, 1985 to 1986

	Model 1[a]		Model 2[b]		Model 3[c]	
Covariate	β	SE (β)	β	SE (β)	β	SE (β)
Maternal race/ethnicity						
White[d]						
Black	−246.56*	2.05	−138.69*	2.26	−120.08*	2.14
American Indian	11.54	18.97	67.71*	18.45	74.43*	17.50
Chinese	−116.40*	7.76	−107.91*	7.71	−115.21*	7.32
Japanese	−177.47*	25.49	−244.46*	24.82	−234.85*	23.55
Filipino	−152.84*	15.88	−176.70*	15.55	−164.37*	14.75
Maternal age, y						
20–34[d]						
≤19			5.93	3.07	16.71*	2.91
≥35			−9.10*	2.88	−0.23	2.74
Marital status						
Unmarried[d]						
Married			102.26*	2.17	85.17*	2.06
Total birth order						
2–3[d]						
1			−58.58*	1.84	−58.93*	1.75
4+			15.25*	2.31	21.64*	2.19
County of residence						
Nonmetropolitan[d]						
Metropolitan			−30.67*	2.07	−38.35*	1.96
Nativity status						
Foreign born[d]						
Native born			−46.44*	2.12	−33.17*	2.01
Infant sex						
Female[d]						
Male			127.41*	1.61	131.70*	1.52
Maternal education			10.95*	0.38	9.18*	0.36
Prenatal visits			20.86*	0.21	15.61*	0.20
Gestational age, wk						
≥37[d]						
<37					−596.05*	2.62
Constant	3402.06*	0.94	2974.41*	5.39	3105.27*	5.14
R^2		0.03		0.08		0.18
df		464 550		464 540		464 539

Note. β = unstandardized regression coefficient.
Source. National Center for Health Statistics, *Linked Birth/Infant Death Data Sets*, 1985 and 1986 birth cohorts.[2,17]
[a]The only covariate included is race/ethnicity.
[b]The covariates included are race/ethnicity, maternal age, marital status, education, nativity, county of residence, birth order, infant sex, and number of prenatal vists.
[c]The covariates included are race/ethnicity, maternal age, marital status, education, nativity, county of residence, birth order, infant sex, number of prenatal visits, and gestational age.
[d]Reference category.
*$P < .01$.

Discussion

This study has examined ethnic differentials in birthweight using resident singleton live-birth records from New York State for 1985 and 1986. The two specific questions addressed were (1) whether the Asian-American groups such as Chinese, Japanese, and Filipinos differed from Whites, Blacks, and American Indians with regard to mean birthweight and risk for low birthweight; and (2) whether the ethnic differences remained after controlling for selected sociodemographic and biological risk factors known to influence birth outcome. The analysis initially revealed important ethnic differences, with Chinese, Japanese, and Filipinos showing significantly lower mean birthweights than White infants. Controlling for such sociodemographic factors as maternal age, education, marital status, nativity, birth order, and prenatal care had a narrowing effect on the relative birthweight differ-

TABLE 3—Logistic Regressions Showing Crude and Adjusted Racial/Ethnic Differentials in Risks for Low Birthweight: New York State, 1985 to 1986

Covariate	Model 1[a] OR	Model 1[a] 95% CI	Model 2[b] OR	Model 2[b] 95% CI	Model 3[c] OR	Model 3[c] 95% CI
Maternal race/ethnicity						
White	1.00		1.00		1.00	
Black	2.48	2.43, 2.54	1.59	1.54, 1.64	1.49	1.44, 1.54
American Indian	1.09	0.82, 1.45	0.82	0.44, 1.54	0.76	0.55, 1.05
Chinese	0.78	0.69, 0.89	0.87	0.75, 1.01	0.93	0.80, 1.08
Japanese	0.79	0.51, 1.22	1.35	0.86, 2.13	1.28	0.80, 2.05
Filipino	1.15	0.91, 1.45	1.54	1.21 1.96	1.45	1.12, 1.87
Maternal age, y						
20–34			1.00		1.00	
≤19			0.87	0.83, 0.90	0.78	0.74, 0.81
≥35			1.28	1.23, 1.34	1.23	1.17, 1.29
Marital status						
Unmarried			1.00		1.00	
Married			0.70	0.68, 0.72	0.76	0.74, 0.79
Total birth order						
2–3			1.00		1.00	
1			1.18	1.14, 1.21	1.21	1.17, 1.25
4+			1.13	1.10, 1.17	1.09	1.05, 1.13
County of residence						
Nonmetropolitan			1.00		1.00	
Metropolitan			0.98	0.94, 1.01	1.09	1.05, 1.13
Nativity status						
Foreign born			1.00		1.00	
Native born			1.42	1.38, 1.47	1.32	1.27, 1.36
Infant sex						
Female			1.00		1.00	
Male			0.83	0.81, 0.85	0.77	0.75, 0.79
Maternal education			0.96	0.96, 0.97	0.97	0.96, 0.97
Prenatal visits			0.89	0.89, 0.89	0.92	0.92, 0.92
Gestational age, wk						
≥37					1.00	
<37					12.44	12.09, 12.78
Model chi-square		5351.91*		13 766.12*		44 237.91*
df		5		15		16
n		499 377		465 310		465 310

Note. OR = odds ratio; CI = confidence interval.
Source. National Center for Health Statistics, *Linked Birth/Infant Death Data Sets*, 1985 and 1986 birth cohorts.[2,17]
[a]The only covariate included is race/ethnicity.
[b]The covariates included are race/ethnicity, maternal age, marital status, education, nativity, county of residence, birth order, infant sex, and number of prenatal visits.
[c]The covariates included are race/ethnicity, maternal age, marital status, education, nativity, county of residence, birth order, infant sex, number of prenatal visits, and gestational age.
*$P < .01$.

ences between Whites and Blacks and Chinese but had a widening impact on the mean birthweight differences between Whites and Japanese, Filipino, and American Indians.

Similarly, observed ethnic differences in rates of low birthweight indicated a significantly lower risk for low birthweight for Chinese and a higher risk for Blacks compared with Whites. Adjustment for the risk factors mentioned above reduced but did not completely eliminate ethnic differences in low birthweight.

Results of this study also point toward the substantial heterogeneity that characterizes the three Asian-American groups with respect to their sociodemographic composition and birth outcome. Of the three groups, Chinese mothers appear to have the most favorable birth outcome although they have lower educational attainment than Japanese and Filipino mothers. Comparison of their standardized low-birthweight rates (after adjusting for risk factors) reveals that Japanese and Filipino mothers were significantly more likely (38% and 56%, respectively) to have a low-birthweight baby than Chinese mothers. In fact, based on ethnic differences in standardized low-birthweight rates, Filipinos appear to resemble Blacks much more closely than they do their Japanese and Chinese counterparts with respect to the risk for low birthweight.

As regards the generalizability of this study, the Asian-American population in New York State may not be entirely representative of the Asian-American population in the United States as a whole. As seen from Table 1, Chinese, Japanese, and Filipino mothers in our sample consist almost entirely of immigrants, with significantly larger proportions of them having been born abroad than is the case at the national level.[11,16] As a result, these mothers differ somewhat from their national counterparts in their sociodemographic composition and therefore in their birth outcomes; for example, Chinese, Japanese, and Filipino mothers in New York State had fewer prenatal visits and were relatively less likely to seek early prenatal care. Furthermore, compared with their national counterparts, Japanese and Filipino mothers in New York State were characterized by higher maternal education and lower rates of teenage, fourth- and higher-order, and out-of-wedlock births.[11,16] Additionally, Chinese mothers in our study sample (which includes New York City) were relatively more likely to be newer immigrants and to be less educated, poor, and employed in secondary labor markets than is the case nationally.

In conclusion, the results of this study indicate important ethnic differences in mean birthweight and risk of low birthweight. But while sociodemographic risk factors considered here account for some of these observed differences, much of the variance in birth outcomes among ethnic groups in general and among Asian Americans in particular remains unexplained. To explain more fully the ethnic differences in mean birthweight and risk of low birthweight will require consideration of additional risk factors not included in the present analysis owing to lack of data. Examples of these risk factors include such socioeconomic variables as income and occupation; behavioral or life-style factors such as maternal cigarette smoking, alcohol consumption,

and other forms of substance abuse; nutritional factors such as maternal and paternal size, gestational weight gain, diet, and breast-feeding; psychosocial[21] and cultural factors such as stress, social and familial support, and cultural attitudes toward maternity; and medical care factors that are more directly related to birth outcomes.[19] Maternal occupation and especially income may contribute substantially to ethnic variations in birth outcomes independent of education, the most commonly used measure of socioeconomic status in the analysis of health outcomes.[22] Maternal physical stature as measured by height and prepregnancy weight is shown to have a profound net effect on birthweight[9,18] and is expected to account for part of the reported birthweight deficit for Asian-American groups considered in this study. However, significant ethnic differentials in both–mean birthweight and risk of low birthweight have been shown to persist even after controlling for ethnic differences in maternal body size and other covariates.[9,23] □

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1990. DHHS publication PHS 91–50212.
2. National Center for Health Statistics. *Linked Birth/Infant Death Data Set: 1986 Birth Cohort, Public Use Data Tape Documentation.* Hyattsville, MD: Public Health Service; 1991.
3. National Center for Health Statistics. Advance report of final natality statistics, 1991. *Month Vital Stat Rep.* 1993;42(3 suppl).
4. Taffel SM. Characteristics of Asian births: United States, 1980. *Month Vital Stat Rep.* 1984;32(10 suppl).
5. Census Bureau releases 1990 census counts on specific racial groups. *Commerce News.* Washington, DC: US Bureau of the Census, US Department of Commerce; June 12, 1991. CB9-215.
6. O'Hare WP, Felt JC. *Asian Americans: America's Fastest Growing Minority Group. Population Trends and Public Policy.* Washington, DC: Population Reference Bureau; 1991.
7. Lin-Fu JS. Asian and Pacific Islander Americans: an overview of demographic characteristics and health care issues. *Asian Am Pacific Islander J Health.* 1993;1:20–35.
8. Lin-Fu JS. Population characteristics and health care needs of Asian Pacific Americans. *Public Health Rep.* 1988;103:18–27.
9. Shiono PH, Klebanoff MA, Graubard BI, Berendes HW, Rhoads GG. Birth weight among women of different ethnic groups. *JAMA.* 1986;255:48–52.
10. Kleinman JC. Infant mortality among racial/ethnic minority groups, 1983–84. *MMWR.* 1990;39(SS-3):31–39.
11. Singh GK, Yu SM. Pregnancy outcomes among Asian Americans. *Asian Am Pacific Islander J Health.* 1993;1:63–78.
12. Park CB, Horiuchi BY. Ethnicity, birth weight, and maternal age in infant mortality: Hawaiian experience. *Am J Human Biol.* 1993;5:101–109.
13. Alexander GR, Baruffi G, Mor JM, Kieffer E. Pregnancy outcomes among whites and Filipinos: a paradoxical birth weight–neonatal mortality relationship. *Am J Human Biol.* 1993;5:203–209.
14. Alexander GR, Baruffi G, Mor JM, Kieffer EC, Hulsey TC. Multiethnic variations in the pregnancy outcomes of military dependents. *Am J Public Health.* 1993;83:1721–1725.
15. Weeks JR, Rumbault RG. Infant mortality among ethnic immigrant groups. *Soc Sci Med.* 1991;33:327–334.
16. Wang X, Strobino DM, Guyer B. Differences in cause specific infant mortality among Chinese, Japanese and White Americans. *Am J Epidemiol.* 1992;135:1382–1393.
17. National Center for Health Statistics. *Linked Birth/Infant Death Data Set: 1985 Birth Cohort, Public Use Data Tape Documentation.* Hyattsville, Md: Public Health Service; 1990.
18. Kramer MS. Determinants of low birthweight: methodological assessment and meta-analysis. *Bull WHO.* 1987;65:663–737.
19. Institute of Medicine. *Preventing Low Birthweight.* Washington, DC: National Academy Press; 1985.
20. SAS Institute, Inc. *SAS/STAT User's Guide: The REG Procedure and LOGISTIC Procedure, Version 6.* 4th ed. Cary, NC: SAS Institute, Inc; 1989;2.
21. Brooks-Gunn J. Stress and support during pregnancy: what do they tell us about low birthweight? In: Berendes H, Kessel S, Yaffe S, eds. *Advances in the Prevention of Low Birth Weight: An International Symposium.* Washington, DC: National Center for Education in Maternal and Child Health; 1991:39–57.
22. Liberatos P, Link BG, Kelsey JL. The measurement of social class in epidemiology. *Epidemiol Rev.* 1988;10:87–121.
23. Dougherty C, Jones A. The determinants of birth weight. *Am J Obstet Gynecol.* 1982;144:190–200.

The Relationship of Prenatal Care and Pregnancy Complications to Birthweight in Winnipeg, Canada

Cameron A. Mustard, ScD, and Noralou P. Roos, PhD

ABSTRACT

Objectives. Prenatal care is commonly understood to have a beneficial impact on birthweight. This study describes socioeconomic differences in utilization of prenatal medical care and birthweight in a population with universal health insurance.

Methods. Measures of prenatal care utilization, incidence of pregnancy complications, and birthweight were obtained from physician reimbursement claims and hospital separation abstracts for 12 646 pregnant women. Maternal socioeconomic status was derived from small-area census data.

Results. Infants born to women in the poorest income quintile had lower birthweights than infants born to wealthier women. Much of the difference was associated with a higher prevalence of complications, smoking, unmarried status, and inadequate prenatal care among low-income women. The difference in birthweight between adequate and less than adequate care groups was small, and the benefit associated with prenatal care was no greater among women with pregnancy complications.

Conclusions. The lower utilization of prenatal care by poorer women accounted for a small proportion of the difference in birthweight. Socioeconomic differences in birthweight are primarily attributable to factors not directly influenced by early prenatal medical care. (*Am J Public Health.* 1994;84:1450–1457)

Introduction

Prenatal care is commonly understood to have a beneficial impact on pregnancy outcome, either through the diagnosis and timely treatment of complications or by contributing to the elimination or reduction of modifiable maternal risk factors. Prenatal care may extend the length of gestation,[1,2] and early initiation of care has been associated with heavier birthweights.[3–6] However, the efficacy and effectiveness of prenatal care as a primary or secondary intervention in the prevention of low birthweight is unresolved. The substantial increase in the proportion of US women initiating care in the first trimester in the past 2 decades has not been accompanied by a significant decline in the incidence of low-birthweight infants.[7,8] Although prenatal care contributes to the identification of complications during pregnancy, the early initiation of regular prenatal care has not been shown to directly reduce the incidence of low birthweight[1,4,6,9–16] or to moderate the incidence of preterm birth.[17–19]

Where prenatal care can be shown to be effective, constraints on access to such care limit the numbers of women who may derive benefit. Canada's universal first-dollar insurance for medical care, although necessary to achieve high population coverage of prenatal services, may be insufficient to achieve equitable access.[20–26] We designed a study with two principal objectives: (1) to describe the utilization of prenatal care in relation to maternal socioeconomic characteristics in a universally insured health care system and (2) to assess the effectiveness of prenatal care, specifically the contribution of prenatal care in mitigating the effects of pregnancy complications on birthweight.

Methods

Population and Study Period

The study population consisted of the 13 349 women resident in metropolitan Winnipeg, Manitoba, who had hospital admissions resulting in live births in the 18-month period between July 1, 1987, and December 31, 1988. Excluded from this population were 288 women (2.2%) whose hospital separation abstract could not be linked to a newborn hospital record and 415 women (3.1%) who were not registered with the Manitoba Health Services Insurance Plan for the complete pregnancy interval. After these exclusions, 12 646 women were retained in the study sample.

Sources of Data

Information on prenatal care was gathered from three computerized administrative databases maintained by the Manitoba Health Services Insurance Plan: hospital separation abstracts; physician claims for reimbursement of ambulatory care, including diagnostic tests performed in a physician's office or a laboratory; and records of the provincial postpartum follow-up nursing program. The three sources of administrative data contain common personal identifiers that permit records to be combined to create a

The authors are with the Manitoba Centre for Health Policy and Evaluation, Department of Community Health Sciences, Faculty of Medicine, University of Manitoba, Winnipeg.

Requests for reprints should be sent to Cameron A. Mustard, ScD, Manitoba Centre for Health Policy and Evaluation, Department of Community Health Sciences, Suite 2008, 351 Taehe Ave, Winnipeg, Manitoba, R2H 2A6, Canada.

This paper was accepted May 25, 1994.

Editor's Note. See related editorial by Wise (p 1374) in this issue.

composite maternal history of prenatal care. These data were supplemented by descriptions of neighborhood socioeconomic characteristics derived from the 1986 Canadian census.

Hospital separation abstracts. Individual separation abstracts are issued for mothers and their newborns. The newborn separation abstract contains information on birthweight, Apgar score at 1 and 5 minutes, gestational age, and a range of *International Classification of Diseases,* Ninth Revision, Clinical Modification[27] (ICD-9-CM) diagnostic codes describing neonatal morbidity. The maternal obstetrical record provides information on age; residential postal code; treaty Indian status; parity; gravidity; ICD-9-CM diagnostic codes defining morbidity associated with pregnancy, labor, and delivery; and a range of ICD-9-CM procedure codes describing the method of delivery and other perinatal obstetrical interventions. The maternal abstract also contains information abstracted from the prenatal care record, which is completed by the primary care physician during pregnancy, including a count of the number of prenatal visits (to a maximum of eight or more), the week of gestation in which care was initiated, and risk assessment scores.

The newborn separation abstract was linked to the mother's record. In cases in which the woman gave birth to more than one baby, only one linked mother-infant pair was retained.

Physician claim files. All physician claims for reimbursement of medical care in Manitoba are submitted to the Manitoba Health Services Insurance Plan. When providing prenatal and postpartum care, a physician may bill on a fee-for-service basis, submitting a claim for each care encounter, or submit a single claim for a fixed-fee global tariff that covers all routine care provided in the prenatal and postpartum periods. In this sample, 86.7% of the pregnancies were supervised by physicians claiming the global tariff.

Postpartum referral files. A referral to a public health nurse is made for all women who give birth, with the objective of completing at least one contact following discharge. The referral record contains information obtained during a predischarge interview with the woman in hospital, including attained education, prenatal care and participation in prenatal education classes, smoking status, and infant feeding practices. Files for 806 (6.4%) of the 12 646 women eligible for study could not be located in the postpartum referral database. These women were assigned the status of nonsmokers and the status of enrollees in prenatal education programs.

TABLE 1—Summary of Measures

	Source
Maternal characteristics	
Mean household income of neighborhood	1986 Canadian census
Self-reported smoking status	Postpartum referral files
Marital status	MHSIP registration files
Registration status under federal Indian Act	MHSIP registration files
Age, in years	MHSIP registration files
Number of previous live births	Hospital separation abstracts
Number of previous pregnancy losses	Hospital separation abstracts
Gestation length, in weeks	Hospital separation abstracts
Pregnancy complications, identified by ICD-9-CM diagnostic code	Physician claim files and hospital separation abstracts
640 Hemorrhage in early pregnancy	
641 Antepartum hemorrhage	
642 Hypertension	
643 Excessive vomiting	
644 Early or threatened labor	
646 Other antenatal complications	
Birthweight, in grams	Hospital separation files
Characteristics of care	
Number of visits during pregnancy	Physician claim files
Time of first contact for pregnancy-related care	Physician claim files
Self-reported maternal participation in prenatal or childbirth education program	Postpartum referral files
Adequacy of care, adjusted for length of gestation[35]	Physician claim files and hospital separation files
Adequate	
Initiation of care in the first trimester and	
9 or more visits if gestation > 35 weeks, or 8 or more visits if 34–35 weeks, or	
7 or more visits if 32–33 weeks, or 6 or more visits if 30–31 weeks, or	
5 or more visits if 26–29 weeks, or 4 or more visits if 22–25 weeks, or	
3 or more visits if 18–21 weeks, or 2 or more visits if 14–17 weeks, or	
1 or more visits if 13 weeks	
Inadequate	
Initiation later than second trimester, or	
1 or fewer visits if gestation 14–29 weeks, or 2 or fewer visits if 30–31 weeks, or	
3 or fewer visits if 32–33 weeks, or 4 or fewer visits if 34 or more weeks	
Intermediate	
All other combinations of care	

Note. MHSIP = Manitoba Health Services Insurance Plan.

Public-use census files. In the absence of individual-level measures of socioeconomic status, data from the 1986 Canadian census, aggregated to the geographic unit of the enumeration area, were used to rank neighborhoods in population quintiles by average household income. This approach has been used in previous research using administrative data.[28–34] In Winnipeg, there are approximately 800 enumeration areas containing residential households, with an average population of 700 people. Using the full six-digit postal code, a total of 71 (0.6%) women could not be linked to census information on mean household income and were assigned the sample average.

TABLE 2—Maternal Demographic Characteristics (%), by Income Quintile, Winnipeg, July 1987 through December 1988

	Income Quintile[a]					Total Sample
	1 (n = 3019)	2 (n = 2144)	3 (n = 2133)	4 (n = 2894)	5 (n = 2456)	(n = 12 646)
Maternal characteristic						
Age, y						
<20	14.1	7.5	5.5	3.8	2.9	7.0
>35	6.5	7.7	7.5	9.5	13.9	9.0
Unmarried	45.3	29.2	23.7	16.5	15.3	26.5
Native treaty status[b]	13.4	3.8	1.7	1.0	0.2	4.4
Parity						
1st birth	44.2	47.4	47.2	43.1	38.9	43.9
4th or later birth	9.9	5.7	5.4	4.3	5.0	6.2
Prior pregnancy loss	25.3	25.9	22.9	23.0	23.0	24.1
Smoked during pregnancy	38.1	33.5	27.9	22.7	15.3	27.6
Complication in pregnancy[c]	29.1	28.0	24.5	23.1	21.5	25.3
Birthweight						
<1500 g	1.2	0.9	0.7	0.7	0.7	0.8
1500–2500 g	5.0	4.6	3.6	3.5	2.8	3.9
Gestation <37 weeks	6.3	6.7	6.2	5.7	4.6	5.9

[a]1 = poorest, 5 = wealthiest.
[b]Registered under the federal Indian Act.
[c]Complications were identified by ICD-9-CM diagnostic codes 640–644, 646 recorded on either physician claims or hospital abstract. See Table 1 for detailed listing.

Measures

Table 1 summarizes the measures of maternal, newborn, and prenatal care characteristics used in this study.

Pregnancy complications. Complications of pregnancy were identified from ICD-9-CM diagnostic codes 640 through 648 on both physician claims and hospital separation abstracts. On the basis of an analysis of the predictive validity of specific diagnoses in relation to birthweight outcome, six diagnostic categories were retained in the analysis (see Table 1). A woman was classified as experiencing a complication if one or more of these diagnoses were identified on either data source.

Frequency of care. Physicians' services for pregnancy care were classified into two categories: office visits and in-office or laboratory diagnostic tests. Office visits were defined as any claim with ICD-9-CM diagnostic codes for normal pregnancy (V22); supervision of high-risk pregnancy (V23); complication related to pregnancy (ICD-9-CM 640–648); other indications for care in pregnancy, labor, and delivery (ICD-9-CM 650–659); or complications occurring in the course of labor and delivery (ICD-9-CM 660–669).

For the 13.3% of women who received care from physicians who billed on a fee-for-service basis, both the timing of initiation of care and the number of episodes of care could be established directly from the individual claims. However, for the majority of women who received care from physicians billing the global tariff, direct enumeration of service claims was not possible. For these women, direct billing of specified in-office or laboratory diagnostic tests was used as a surrogate measure of the number of episodes of care. The selected tests were glucose monitoring, hemoglobin, urinalysis, Pap smear, and hormonal pregnancy test. The first three tests are the standard monitoring tests in pregnancy.

A comprehensive ambulatory health care history was compiled for each woman in the sample, capturing all instances of use of ambulatory care services in the period from 365 days before birth to 90 days postpartum. Of the 497 062 claims abstracted, approximately 54% were associated with the provision of prenatal or postpartum care and were retained.

Adequacy of care. Length of pregnancy in days was calculated individually for each woman from the measure of gestation period on the hospital record. For the 41 cases (0.3%) with a missing or out-of-range gestation period, a gestation period of 280 days was assigned. A summary measure of prenatal care, defining the adequacy of care relative to an ideal standard, was constructed from measures of the timing of initiation of care, frequency of visits, and the length of gestation.[35] This index, which classifies care as adequate, intermediate, or inadequate, has been used in other settings to adjust for the effects of shortened gestation on the use of prenatal care.[6,10,36]

A series of analyses was performed to examine the agreement between physician claim records and the hospital abstract on measures of utilization of care. A detailed description of these analyses is presented elsewhere.[37] In general, physician claims were found to yield measures of care comparable to those of the hospital abstract. However, the hospital abstract was missing information on the initiation of care for 12.3% of the women and on the number of prenatal visits for 9.2% of the women. In subsequent analyses reported here, measures of use derived from physician claims have been applied exclusively, in contrast to the alternative approach of using physician claim measures only to substitute for missing utilization measures on the hospital abstracts.

Analytic Methods

To assess the effectiveness of prenatal care, we compared pregnancy outcomes across groups of women who received different amounts of care or who initiated care at different points in their pregnancies. In this observational design, there are three methodological obstacles for which no ideal analytic solution exists.[33] The most fundamental problem in comparing outcomes across groups of women with different patterns of utilization is that women who give birth prematurely will most certainly be shown to use less care than women whose pregnancies go to term. A second problem, with an opposite effect on utilization, is confounding by indication: women who experience complications early in pregnancy will be

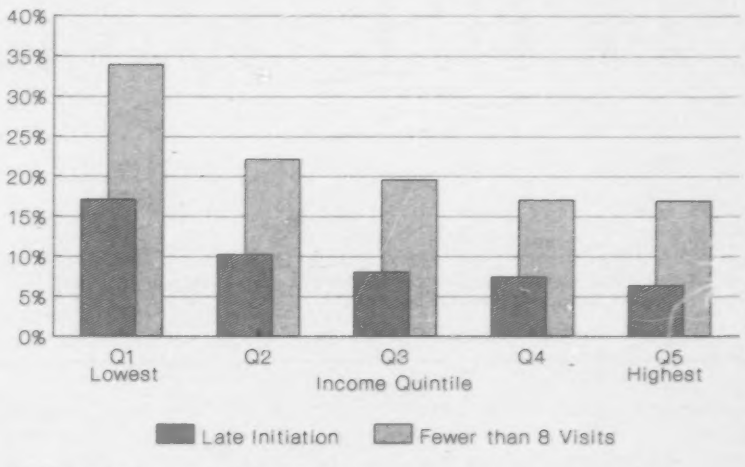
TABLE 3—Characteristics of Prenatal Care, by Income Quintile, Winnipeg, July 1987 through December 1988

	Income Quintile[a]						
	1 (n = 3019)	2 (n = 2144)	3 (n = 2133)	4 (n = 2894)	5 (n = 2456)	Total Sample (n = 12 646)	P
No. of visits							
<5, %	14.8	7.4	6.7	5.6	5.0	8.2	<.001[b]
5–8, %	19.1	14.7	12.8	11.4	11.9	14.1	
>13, %	21.9	29.7	32.5	32.3	32.5	29.3	
Average no.	10.0	11.2	11.4	11.6	11.7	11.2	<.001[c]
Initiation of care							
Mean week of initiation	8.9	7.7	7.3	7.2	7.2	7.7	<.001[c]
Late initiation							
Second trimester, %	13.0	8.4	6.3	6.0	4.9	7.8	<.001[b]
Third trimester, %	4.1	1.8	1.7	1.4	1.4	2.2	
Adequacy of care							
Adequate, %	61.1	74.0	77.9	80.0	81.1	74.3	<.001[b]
Intermediate, %	22.9	18.0	14.5	13.7	13.3	16.7	
Inadequate, %	15.9	7.9	7.5	6.0	5.6	8.9	

[a]1 = poorest, 5 = wealthiest.
[b]Chi-square test of association, testing for differences across income strata.
[c]Analysis of variance, testing for differences across income strata.

expected to seek medical care earlier and more often than women with an uncomplicated first trimester. The final obstacle to valid inferences about the effect of care on outcome in an observational study is selection bias. The analytic approach of this study was designed to minimize these problems. To reduce the chance of bias due to preterm delivery, we adopted an index of the adequacy of prenatal care that classifies the use of care according to the length of gestation; however, this classification precludes examining the potential effect of prenatal care on pregnancy duration in the absence of fetal weight gain. We also emphasized the ascertainment of pregnancy complications, which permitted statistical adjustment for the influence of complications on use of care. No adjustment for errors of inference due to selection effects is possible in an observational research design.

A chi-square test of association or an *F*-test statistic derived from an analysis of variance was used to examine differences between women in demographic risk factors, prenatal care utilization, and the incidence of complications. The relationship of measures of prenatal care, maternal characteristics, and pregnancy complications to birthweight was examined in two ways: (1) birthweight in grams was regressed on the set of independent variables in a multivariate linear regression analysis and (2) birthweight was classified as a dichotomous dependent variable, with weight less than 2500 g defined as low birthweight in a logistic regression analysis. As the main results of these two models were similar, only the linear regression analyses are reported here.



Note. Late initiation = more than 13 weeks gestation.

FIGURE 1—Initiation of prenatal care and number of visits, by income quintile.

Results

Characteristics of the population, stratified by income quintile, are reported in Table 2. First births accounted for 43.9% of the births. A total of 7% of the women in the study population were younger than 20 years of age. Smoking during pregnancy was reported by 27.6% of the women, 26.5% were unmarried at the time of delivery, 24.1% had experienced a prior pregnancy loss, and a complication in the prenatal period was detected in 25.3% of the pregnancies. Overall, 4.7% of the infants weighed less than 2500 g at birth, and 5.9% of the births occurred before the 37th week of gestation.

TABLE 4—Unadjusted Mean Birthweight, by Complication Status and Adequacy of Prenatal Care, Winnipeg, July 1987 through December 1988

	n	Mean Birthweight, g	95% Confidence Interval
No complication	9450	3473	
Complication	3196	3229	
No complication			
Adequate care	7125	3491	...
Less than adequate care	2325	3418	3392, 3444
Complication			
Adequate care	2437	3247	3222, 3272
Less than adequate care	759	3175	3134, 3217

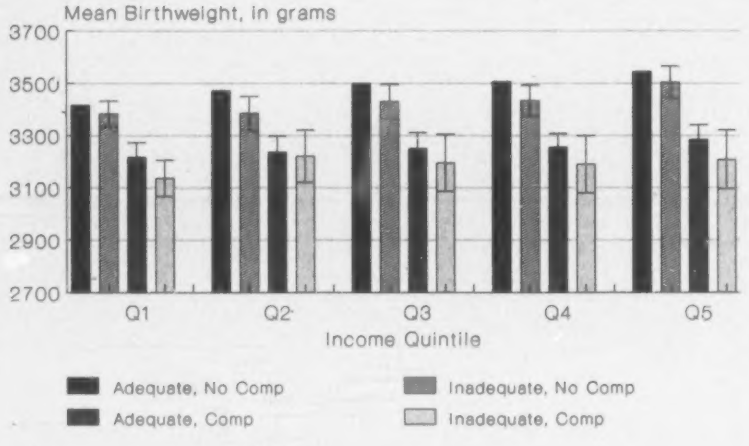


FIGURE 2—Mean birthweight, by income quintile, pregnancy complications (Comp), and prenatal care.

Maternal characteristics were distributed unequally across the five neighborhood income categories. Women residing in the poorest 20% of neighborhoods (quintile 1), compared with women residing in the wealthiest 20% of neighborhoods (quintile 5), were more than twice as likely to report having smoked during pregnancy, were three times as likely to be unmarried, and were four times as likely to be under the age of 20. The crude incidence of low birthweight among women in quintile 1 was 62 per 1000 live births, almost twice the rate observed among women in quintile 5 (35/1000).

Women in the sample had an average of 11.2 visits to prenatal care professionals during their pregnancies (Table 3). A total of 8.2% (1037 women) had fewer than 5 visits, and 14.1% had between 5 and 8 visits. Care was initiated, on average, during the 7th week of gestation; 90% of the women initiated care by the 13th week. In only 2.2% of the pregnancies was the first visit delayed until the third trimester. When utilization was expressed in terms of adequacy of care, 74.3% of the women received adequate care and 8.9% received inadequate care; inadequate care was defined as initiation later than the second trimester (27 weeks) or, in the case of earlier initiation, an insufficient number of prenatal care contacts.

Only marginal differences in utilization were observed among the 60% of women in the three highest income quintiles. However, the 20% of women in the lowest income quintile had consistently and substantially poorer utilization than those in the median- and high-income groups. Poor women had an average of 1.5 fewer visits during their pregnancies and first sought care an average of 1.5 weeks later. A total of 17% of the poorest women initiated care later than the first trimester, compared with less than 7% of the women in the median to high quintiles (Figure 1).

Among uncomplicated pregnancies, infants born to women who received less than adequate care were, on average, 73 g lighter than infants of women receiving adequate prenatal care ($P < .001$, 95% confidence interval [CI] = 47 99) (Table 4). After adjustment for maternal characteristics and income, this difference was reduced to an estimated 55.6 g ($P < .001$, data not shown). In these analyses, prenatal care was dichotomized as adequate or less than adequate. The latter category combines both intermediate and inadequate care.

As expected, the average birthweight of infants born to the 25.2% of women who experienced complications in pregnancy was substantially lower than that of infants born to women with uncomplicated pregnancies (a difference of 244 g; $P < .001$, 95% CI = 185, 301) (Table 4). The complications measured in this study are prevalent in early gestation and are strongly predictive of premature birth.

Among women with complicated pregnancies, only a slight difference in birthweight (72 g; $P < .001$, 95% CI = 31, 114) was observed between the adequate care and less than adequate care groups. The estimated positive contribution of care to birthweight was no greater in complicated pregnancies than in pregnancies without noted complications. As shown in Figure 2, this weak relationship was consistent across all income groups.

In a bivariate analysis, mean birthweight decreased with declining income. Infants born to women in the poorest quintile were, on average, 140 g lighter than infants born to women in the wealthiest quintile (Table 5). The magnitude of birthweight differences associated with income was reduced when maternal factors were included in the analytic model, reflecting the concentration of maternal risk factors such as smoking and single marital status among women living in poor neighborhoods. These income-related differences were further reduced following the addition of variables measuring pregnancy complications and use of prenatal care.

These analyses were repeated after stratifying the population into two groups according to the presence or absence of recorded complications in pregnancy. The

results were similar to those of the unstratified analysis.

Discussion

This study has a number of strengths compared with previous studies of the effectiveness of early and frequent prenatal care. Information on prenatal care utilization, maternal smoking status, the incidence of pregnancy complications, and an indirect measure of socioeconomic status were available for a virtually complete population of urban women giving birth between July 1987 and December 1988.

The study provided a direct test of the hypothesis that prenatal care has the greatest impact on complicated pregnancies. Pregnancy complications were detected in 25.2% of the women in this study, an incidence higher than has typically been reported in other studies. However, the literature describing the incidence of complications in complete populations is poorly standardized, with wide variations in measurement methods, sources of data, and inclusion criteria.[6,10,14,30,38–43] In general, however, complication-specific rates observed in the Winnipeg population are within the ranges reported in other developed country settings. A British study reported a rate of antepartum hemorrhage of 8.8%,[43] which compares well with the rate of 6.9% observed in the Winnipeg population. Incidence rates of pre-eclampsia and pregnancy-induced hypertension have been reported in the range of 6.5 to 6.2 per 100 births[39,43]; these rates agree with the rate of 6.5 per 100 observed in this study.

Utilization of Prenatal Care

The adequacy of the care received by Winnipeg women under a system of universal insurance that includes comprehensive prenatal medical services compares favorably with rates observed in other North American populations.[8] As measured by trimester of initiation of care, access to prenatal care was better for the Winnipeg women than has been reported for typical US samples, and these differences are more pronounced among women of lower socioeconomic status. A total of 82.9% of the poor women in the Winnipeg sample initiated care by the 13th week of pregnancy, compared with 69% of poor women in an urban sample in Washington State[44] and 70% of low-income women not eligible for Medicaid and 62% of Medicaid-insured women in southern California.[14] Despite this favorable comparison, it is clear that universal health insurance does not remove all barriers to access. The Winnipeg women in the lowest income quintile had consistently and substantially poorer utilization of care than did median- and high-income women.

TABLE 5—Multivariate Linear Regression Analysis of the Relationship of Maternal Characteristics, Pregnancy Complications, and Adequacy of Prenatal Care to Birthweight, Winnipeg, July 1987 through December 1988

	Income Only		Income, Maternal Factors		Income, Maternal Factors, Complications, Adequacy of Care	
R^2	0.7		4.1		7.2	
Intercept, g	3480.0		3703.9		3785.6	
	Difference	***P***	**Difference**	***P***	**Difference**	***P***
Income quintile						
1	−139.1	<.001	−107.6	<.001	−88.8	<.001
2	−91.1	<.001	−55.4	<.001	−41.4	<.01
3	−57.4	<.001	−32.2	NS	−25.8	NS
4	−44.8	<.01	−32.5	<.05	−29.3	<.05
5	Ref		Ref		Ref	
Treaty status						
Non-Native			−139.4	<.001	−163.3	<.001
Native			Ref		Ref	
Maternal age, y						
<20			10.8	NS	23.5	NS
20–34			Ref		Ref	
>34			−62.8	<.001	−57.9	<.001
Parity						
1st birth			−183.5	<.001	−167.5	<.001
2nd–4th birth			Ref		Ref	
5th or later birth			17.3	NS	41.8	NS
Marital status						
Unmarried			−4.3	NS	−9.1	NS
Married			Ref		Ref	
Smoking status						
Smoker			−168.9	<.001	−163.8	<.001
Nonsmoker			Ref		Ref	
Prior pregnancy loss						
Yes			−31.2	<.01	−22.0	<.01
No			Ref		Ref	
Prenatal education						
Yes			38.0	<.05	32.3	NS
No			Ref		Ref	
Pregnancy complication						
Yes					−228.9	<.01
No					Ref	
Adequate care						
No					−58.6	<.001
Yes					Ref	
Inadequate Care × Complication					−4.1	NS

Note. Difference = estimated mean difference in grams, relative to reference groups (Ref); NS = not significant.

Effectiveness of Prenatal Care

After adjustment for maternal characteristics and early complications of pregnancy, infants born to women receiving less than adequate care were 58 g lighter than infants born to women receiving adequate care. The size of this benefit is much smaller than has been reported in some recent studies, in which adequate

care has been associated with birthweights 157 to 197 g higher than the birthweights of infants born to women with inadequate care.[6,10,41]

The smaller effect of care in the present study may have been due to a number of factors. The Winnipeg cohort may have been healthier than other samples, with a lower rate of unfavorable birthweights. The overall incidence of low birthweight in the Winnipeg sample during the study interval was 47 per 1000 births. A similarly small effect of care on birthweight was reported among women enrolled in a northern California health maintenance organization (HMO), where the rate of low birthweight was 42 per 1000 births.[45] In this HMO population, in which 83% of women initiated care in the first trimester, the estimated contribution of early initiation of care to birthweight, after maternal factors were adjusted for, was 29.9 g. This pattern suggests that the beneficial effects of prenatal care on birthweight become increasingly subtle as the reproductive health of a population improves, and it argues for programs focused on specific populations at high risk.

A surprising and unsettling finding of this study is the absence of a substantial effect of prenatal care on birthweight in complicated pregnancies. The premise of prenatal care is that regular surveillance will enable caregivers to detect the onset of complications and initiate therapeutic responses to mitigate their negative impacts. There is no biological model to suggest any benefit of prenatal medical care in an uncomplicated normal pregnancy. In the management of complications, prenatal care will certainly influence pregnancy outcome in ways not measured by birthweight. At the same time, the absence of a significant impact of care on birthweight in complicated pregnancies in this study provides further evidence of the limited effectiveness of current clinical approaches to the prevention of preterm birth.

The failure to demonstrate a substantial beneficial effect of early prenatal care on birthweight in complicated pregnancies may be due to errors in our analytic approach. Most of the complications examined in this study are strongly predictive of preterm birth. Identification and successful treatment of these complications in the course of prenatal care would result in longer gestation and increased average birthweight. However, the analytic approach used in this study would not be sensitive to interventions that extended the duration of pregnancy, thereby contributing to fetal maturity at birth, but that did not enhance fetal weight gain. It is also possible that the decision to combine a group of discrete complications into a single measure obscured successful complication-specific interventions. Finally, we did not test the hypothesis that prenatal medical care can actually prevent the onset of complications. To the extent that this is a plausible expectation, this study has underestimated the contribution of care by failing to estimate averted morbidity in the group without complications.

In this study, we emphasized the description of socioeconomic differences in the use of prenatal care and in pregnancy outcome. After adjusting for maternal demographic factors, smoking status, complications, and the use of prenatal care, we found a residual difference in average birthweight of 88.8 g between the wealthiest and poorest women in the sample. This unexplained birthweight deficit is larger than can be attributed to the combined effects of complications, smoking, single marital status, and poorer utilization of prenatal care among poor women. Compared with middle-income and wealthy women, poor women experience greater nutritional deficits in pregnancy, may be more likely to be employed in physically demanding occupations, may be exposed more frequently to vaginal infections, and in general have fewer resources with which to respond to the demands of pregnancy.[17,19,46–49] Many of these factors are largely beyond the reach of prenatal medical care.

The findings of this study indicate that the impact of pregnancy complications on birthweight will not be substantially mitigated by improving the utilization of routine prenatal care. After 2 decades of observational studies, the evidence for a beneficial effect of early and frequent prenatal care on birthweight remains ambiguous. To resolve this uncertainty, it has been suggested that researchers conduct randomized trials of specific questions about the organization and delivery of prenatal care.[9,19,50] Without evidence from such definitive studies, the findings of this study do not support the proposition that the route to moderating socioeconomic inequities in birthweight is through improved access to conventional prenatal care. □

References

1. Institute of Medicine. *Preventing Low Birthweight.* Washington, DC: National Academy Press; 1985.
2. Papiernik E, Bouyer J, Dreyfus J. Prevention of preterm births: a perinatal study in Haguenau, France. *Pediatrics.* 1985;76:154–158.
3. Eisner V, Brazie JV, Pratt MW. The risk of low birthweight. *Am J Public Health.* 1979; 69:887–893.
4. Greenberg RS. The impact of prenatal care in different social groups. *Am J Obstet Gynecol.* 1983;145:797–801.
5. Gortmaker SL. The effects of prenatal care upon the health of the newborn. *Am J Public Health.* 1979;69:653–660.
6. Showstack JA, Budetti PP, Minkler D. Factors associated with birthweight: an exploration of the roles of prenatal care and length of gestation. *Am J Public Health.* 1984;74:1003–1008.
7. McCormick MC. The contribution of low birth weight to infant mortality and childhood morbidity. *N Engl J Med.* 1985;312:82–90.
8. Ingram DD, Makuc D, Kleinman JC. National and state trends in use of prenatal care, 1970–83. *Am J Public Health.* 1986;76: 415–423.
9. Murray JL, Bernfield M. The differential effect of prenatal care on the incidence of low birth weight among blacks and whites in a prepaid health care plan. *N Engl J Med.* 1988;319:1385–1391.
10. Quick JD, Greenlick MR, Roghmann KJ. Prenatal care and pregnancy outcome in an HMO and general population: a multivariate cohort analysis. *Am J Public Health.* 1981;71:381–390.
11. Schwartz S, Vinyard JH. Prenatal care and prematurity. *Public Health Rep.* 1965;80: 237–248.
12. Peoples MD, Grimson RC, Daughtry GL. Evaluation of the effects of the North Carolina Improved Pregnancy Outcome Project: implications for state-level decision-making. *Am J Public Health.* 1984;74:549–554.
13. Carey TS, Weis K, Homer C. Prepaid versus traditional Medicaid plans: lack of effect on pregnancy outcomes and prenatal care. *Health Serv Res.* 1991;26:165–181.
14. Howell M, Herz EJ, Wang R, Hirsch MB. A comparison of Medicaid and non-Medicaid obstetrical care in California. *Health Care Financing Rev.* 1991;12:1–15.
15. Peoples MD, Siegel E. Measuring the impact of programs for mothers and infants on prenatal care and low birth weight: the value of refined analyses. *Med Care.* 1983;21:586–605.
16. Piper JM, Ray WA, Griffin MR. Effects of Medicaid eligibility expansion on prenatal care and pregnancy outcome in Tennessee. *JAMA.* 1990;264:2219–2223.
17. Lumley J. The prevention of preterm birth: unresolved problems and work in progress. *Aust Paediatr J.* 1988;24:101–111.
18. Konte JM, Creasy RK, Laros RK. California North Coast Preterm Birth Prevention Project. *Obstet Gynecol.* 1988;71:727–730.
19. Kramer MS. Determinants of low birth weight: methodological assessment and

meta-analysis. *Bull World Health Organ.* 1987;65:663–737.
20. Guyer B. Medicaid and prenatal care: necessary but not sufficient. *JAMA.* 1990; 264:2264–2265.
21. Hall MH, Chng PK, MacGillivray I. Is routine antenatal care worth while? *Lancet.* 1980;2:78–80.
22. Oakley A, Rajan L, Grant A. Social support and pregnancy outcome. *Br J Obstet Gynaecol.* 1990;97:155–162.
23. Cooney JP. What determines the start of prenatal care? *Med Care.* 1985;23:986–997.
24. Miller CL, Margolis LH, Schwethelm B, Smith S. Barriers to implementation of a prenatal care program for low income women. *Am J Public Health.* 1989;79:62–64.
25. Hansell MJ. Sociodemographic factors and the quality of prenatal care. *Am J Public Health.* 1991;81:1023–1028.
26. Chng PK, Hall MH, MacGillivray I. An audit of antenatal care: the value of the first antenatal visit. *BMJ.* 1980;281:1184–1186.
27. *International Classification of Diseases,* Ninth Revision, Clinical Modification. Hyattsville, Md: National Center for Health Statistics; 1980. DHHS publication PHS 80–1260.
28. McLaughlin CG, Normolle DP, Wolfe RA, et al. Small-area variation in hospital discharge rates: do socio-economic variables matter? *Med Care.* 1989;27:507–521.
29. Anderson G, Grumbach K, Luft H, Roos LL, Mustard C. Use of coronary artery bypass surgery in the United States and Canada: influence of age and income. *JAMA.* 1993;269:1661–1666.
30. Wilkins R. Use of postal codes and addresses in the analysis of health data. *Health Rep.* 1993;5:157–177.
31. Wissow LS, Gittelsohn AM, Szklo M, Starfield B, Mussman M. Poverty, race, and hospitalization for childhood asthma. *Am J Public Health.* 1988;78:777–782.
32. Satin MS, Monetti CH. Census tract predictors of physical, psychological, and social functioning for needs assessment. *Health Serv Res.* 1985;20:341–358.
33. Logue EE, Jarjoura D. Modeling heart disease mortality with census tract rates and social class mixtures. *Soc Sci Med.* 1990;31:545–550.
34. Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *Am J Public Health.* 1992;82:703–710.
35. Kessner DM, Singer J, Kalk CW. Infant death: an analysis by maternal risk and health care. In: Kessner DM, ed. *Contrasts in Health Status.* Washington, DC: Institute of Medicine; 1973: vol 1.
36. Alexander GR, Tompkins ME, Petersen DJ, Weiss J. Source of bias in prenatal care utilization indices: implications for evaluating the Medicaid expansion. *Am J Public Health.* 1991;81:1013–1016.
37. Mustard CA. *The Utilization of Prenatal Care and Relationship to Birthweight Outcome in Winnipeg, 1987–88.* Winnipeg, Manitoba, Canada: Manitoba Centre for Health Policy and Evaluation. Report 01.93.
38. Stockbauer JW. WIC prenatal participation and its relation to pregnancy outcomes in Missouri: a second look. *Am J Public Health.* 1987;77:813–818.
39. Leppert PC, Namerow PB, Barker D. Pregnancy outcomes among adolescent and older women receiving comprehensive prenatal care. *J Adolesc Health Care.* 1986;7:112–117.
40. Rosenberg SN, Albertsen PC, Jones EE, Roberts RS. Complications of labor and delivery following uncomplicated pregnancy. *Med Care.* 1981;19:68–79.
41. Scholl TO, Miller LK, Salmon RW, Cofsky MC, Shearer J. Prenatal care adequacy and the outcome of adolescent pregnancy: effects on weight gain, preterm delivery and birth weight. *Obstet Gynecol.* 1987;69: 312–316.
42. Alexander GR, Cornely DA. Prenatal care utilization: its measurement and relationship to pregnancy outcome. *Am J Prev Med.* 1987;3:243–253.
43. Thomas PT, Golding J, Peters TJ. Delayed antenatal care: does it effect pregnancy outcome? *Soc Sci Med.* 1991;32:715–723.
44. Fisher ES, LoGerfo JP, Daling JR. Prenatal care and pregnancy outcomes during the recession: the Washington State experience. *Am J Public Health.* 1985;75:866–869.
45. Shiono PH, Klebanoff MA, Graubard BI, Berendes HW, Rhoads GG. Birth weight among women of different ethnic groups. *JAMA.* 1986;255:48–52.
46. Mustard CA. *Maternal Demographic Risk Factors and the Incidence of Low Birthweight, Manitoba, 1979–89.* Winnipeg, Manitoba, Canada: Manitoba Centre for Health Policy and Evaluation; November 4, 1991. Report 04.91.
47. Lieberman E, Ryan KJ, Monson RR, Schoenbaum SC. Risk factors accounting for racial differences in the rate of premature birth. *N Engl J Med.* 1987;317:743–748.
48. Savitz DA, Whelan EA, Kleckner RC. Effects of parent's occupational exposures on risk of stillbirth, preterm and small-for-gestational-age infants. *Am J Epidemiol.* 1989;129:1201–1218.
49. Johns Hopkins Study of Cervicitis. Association of *Chlamydia trachomatis* and *Mycoplasma hominis* with intrauterine growth retardation and preterm delivery. *Am J Epidemiol.* 1989;129:1247–1257.
50. Tyson J, Guzick D, Rosenfeld CR, et al. Prenatal care evaluation and cohort analyses. *Pediatrics.* 1990;85:195–204.

Is Breast-Feeding in Infancy Associated with Adult Longevity?

ABSTRACT

Objectives. The purpose of the study was to determine whether breast-feeding is associated with increased longevity or cause-specific survival.

Methods. Teachers throughout California identified intellectually gifted children as part of a prospective study begun in the 1920s by Lewis Terman. Information on breast-feeding was available on 1170 subjects, who have been followed for more than 65 years.

Results. Survival analysis (Cox proportional hazards model) indicated that breast-feeding was associated with increased longevity, even after adjustment for age at baseline, birthweight, infant health, and childhood socioeconomic status, but only among men, and the association was not significant ($P = .15$). Neither cardiovascular disease nor cancer survival was significantly associated with duration of breast-feeding for either sex. Survival from deaths due to injuries was positively associated with breast-feeding after adjustment ($P = .03$) and demonstrated a clear gradient with duration, but only among men.

Conclusions. Overall, the present study does not provide strong evidence that breast-feeding is associated with adult longevity. The reduced risk of death from injury may reflect chance, in that the association was significant only for men, or it may reflect psychosocial correlates of breast-feeding practices. (*Am J Public Health.* 1994;84:1458–1462)

Deborah L. Wingard, PhD, Michael H. Criqui, MD, MPH, Sharon L. Edelstein, ScM, Joan Tucker, PhD, Carol Tomlinson-Keasey, PhD, Joseph E. Schwartz, PhD, and Howard S. Friedman, PhD

Introduction

The immediate benefits of breast-feeding are well known.[1,2] Breast-feeding increases survival in poor countries and reduces the risk of infant gastrointestinal infections, lower respiratory tract infections, otitis media, bacteremia, and meningitis in all settings. This benefit is most evident in the first 6 months of life, but it persists into the child's second year. Studies of long-term effects have demonstrated that breast-feeding is associated with a lower risk of childhood cancer, especially lymphoma[3]; insulin-dependent diabetes mellitus[4]; and Crohn's disease and ulcerative colitis.[5,6]

In a previous study of mortality rates, men who had been breast-fed as infants and weaned before their first birthday, or who had been both breast- and bottle-fed, had lower ischemic heart disease mortality rates than men who were either exclusively bottle-fed or breast-fed and not weaned before their first birthday.[7–9] In addition, as adults the former group also had lower serum total cholesterol levels, lower low-density lipoprotein (LDL) levels, and lower ratios of LDL to high-density lipoprotein (HDL).[9] Studies in children and young adults indicate that long-term differences in serum cholesterol levels between breast-fed and bottle-fed infants are small and inconsistent.[10] In an experimental study of baboons, animals that had been breast-fed had higher very low-density lipoprotein levels and higher LDL to HDL ratios until adulthood, and more extensive arterial fatty acid streaks as adults, than did animals that had been bottle-fed.[11]

In the present study we examine whether breast-feeding is associated with increased longevity and cause-specific survival in a prospective population-based study, the Terman Life Cycle Study.[12] Data are available on more than 1000 boys and girls born between 1904 and 1915 who have been followed for mortality through 1986.

Methods

Sample

The Terman Life Cycle Study was initiated by Lewis Terman in the 1920s and continued first under the direction of Robert Sears and then Albert Hastorf. Teachers throughout California identified the three brightest children and the youngest child in their classrooms for testing. This yielded a sample of 1528 children (856 boys and 672 girls) who were identified as gifted. This sample is fairly homogeneous on intelligence (bright), race (White), and social class (mostly middle).[13]

The present analyses were restricted to the 1373 subjects who were born

Deborah L. Wingard, Michael H. Criqui, and Sharon L. Edelstein are with the Department of Family and Preventive Medicine, University of California, San Diego. Joan Tucker is with the Department of Psychology, Brandeis University, Waltham, Mass. Carol Tomlinson-Keasey is with the Department of Psychology, University of California, Davis. Joseph E. Schwartz is with the Department of Psychiatry and Behavioral Sciences, State University of New York at Stony Brook. Howard S. Friedman is with the Department of Psychology, University of California, Riverside.

Requests for reprints should be sent to Deborah L. Wingard, PhD, Department of Family and Preventive Medicine 0607, University of California, San Diego, La Jolla, CA 92093-0607.

This paper was accepted November 12, 1993.

TABLE 1—Percentage Distribution of Breast-Feeding Duration, by Infant's Sex: Subjects in the Terman Life Cycle Study

Breast-Feeding Duration	Boys (n = 647)	Girls (n = 523)
No breast-feeding	10.7	8.8
1–5 mo	22.6	25.2
6–11 mo	34.6	34.6
12+ mo	32.2	31.4
Mean (range)	8.2 (0–24)	8.0 (0–36)

TABLE 2—Distribution of Potentially Confounding Variables, by Breast-Feeding Duration and Sex: Subjects in the Terman Life Cycle Study

Breast-Feeding Duration	Baseline Age, y		Birthweight, lb		Socioeconomic Status[a]		Poor or Fair Infant Health	
	Mean	SD	Mean	SD	Mean	SD	No.	%
Boys (n = 639)								
No breast-feeding	11.5	2.8	8.3	1.6	0.01	0.85	22	31.9
1–5 mo	11.1	2.8	8.2	1.4	0.02	0.80	41	28.9
6–11 mo	11.5	3.1	8.4	1.5	0.08	0.87	36	16.1
12+ mo	12.2*	3.0	8.5	1.5	−0.21	0.85	23	11.4***
Girls (n = 519)								
No breast-feeding	11.1	2.8	7.4	1.5	0.03	0.80	9	20.0
1–5 mo	11.0	2.9	7.8	1.1	0.07	0.87	23	17.8
6–11 mo	10.9	2.8	8.0	1.6	0.09	0.80	13	7.3
12+ mo	11.8	2.8	8.1**	1.5	−0.37**	0.85	11	6.8***

Note. Data were missing on birthweight for 15 boys and 15 girls, on socioeconomic status for 3 boys and 3 girls, and on infant health for 11 boys and 9 girls. *P* values were determined by testing for linear trend based on a linear contrast in an analysis of variance (means) or Mantel-Haenszel chi-square test for trend (percentages).

[a]Parents' education and occupation were standardized to a mean of 0 and a variance of 1 and then averaged. Postive values indicate a higher socioeconomic status, and negative values indicate a lower socioeconomic status.

*$P < .05$; **$P < .01$; ***$P < .001$.

between 1904 and 1915, inclusive (average birth year 1910), to avoid possible cohort effects and to be comparable with our other publications on this data set. Of these participants, 25 dropped out of the Terman study before 1930 and are excluded from the analysis. The 104 subjects lost to follow-up after 1930 are included in the survival analyses up to their date of last contact. A total of 178 subjects were missing information on breast-feeding, resulting in a final sample size of 1170 (647 males, 523 females). Mean age at baseline (approximately 1922) was 11.7 years for the boys and 11.2 years for the girls.

Variables

In the first assessment, parents answered the question, "If child was breast-fed, from what age to what age?" The answers were recorded in months. For the present analyses, the number of months breast-fed was recoded into four roughly equal categories that represent natural cutpoints frequently used in other breast-feeding research (0, 1–5, 6–11, and 12–36 months). At this time, parents also reported their child's year of birth, birthweight, the father's occupation, and the highest grade of school each parent had completed. Parents' education and occupation were standardized to a mean of 0 and a variance of 1 and then averaged to obtain an estimate of the family's socioeconomic status. Positive values indicate a higher social class, and negative values indicate a lower social class; however, the absolute range was narrow and most subjects were in the middle social class.

Parents were also asked to describe their child's health during his or her first year of life. On the basis of these descriptions, the Terman research group rated the child's infant health on a 5-point scale (1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = very good). For the present analyses, this variable was dichotomized by collapsing ratings 1 through 3 and ratings 4 and 5.

At the first visit, one or both of the subject's parents (if one, usually the mother) and the subject's teacher were asked to rate the subject on 25 trait dimensions. Through factor analysis, six personality dimensions have been identified: conscientiousness, cheerfulness, high motivation/self-esteem, sociability, high energy, and permanency of moods. The first two dimensions are significantly associated with adult longevity.[14]

Surviving subjects were contacted every 5 to 10 years. Death certificates were obtained for the deceased and coded for underlying cause of death by a certified nosologist, using the ninth revision of the *International Classification of Diseases*.[15] Cardiovascular disease deaths were those coded 400–448; coronary heart disease included codes 410–414; cancer included codes 140–239; and injury (both intentional and unintentional) included codes 800–998. For 100 of the 449 deceased subjects, no death certificates were obtained. Physicians' review of information from next of kin allowed classification of 65 of these deaths into broad categories of causes, including cardiovascular disease, coronary heart disease, cancer, and injury. Cause-specific analyses including and excluding these latter deaths indicated no difference in findings; therefore the cause-specific analyses presented use all available information.

Analyses

The sex-specific association between breast-feeding duration and potentially confounding variables was assessed by means of linear contrasts in analysis-of-variance models (for age at baseline, birthweight, and childhood socioeconomic status) and the Mantel-Haenszel chi-square test for trend (for infant health). The association between breast-feeding duration and infant health was also assessed with logistic regression analysis, with and without control for age at baseline, birthweight, and childhood socioeconomic status. The sex-specific association between breast-feeding and personality traits, as a possible mechanism for any association with longevity, was assessed with analysis of variance.

Breast-feeding and breast-feeding duration were used to predict longevity from baseline until 1986 by means of survival analysis (Cox proportional hazards model).[16] The group with any breast-feeding and those in each category of breast-feeding duration were compared with the non–breast-feeding group. The analyses, conducted for all causes as well as specific causes of death, were conducted separately for males and females,

TABLE 3—Total and Cause-Specific Mortality Rates (%), by Breast-Feeding Duration and Sex: Subjects in the Terman Life Cycle Study

Breast-Feeding Duration	All Causes		Cardiovascular Disease		Cancer		Injury		Other	
	Males (n = 284)	Females (n = 165)	Males (n = 90)	Females (n = 37)	Males (n = 88)	Females (n = 52)	Males (n = 42)	Females (n = 17)	Males (n = 43)	Females (n = 45)
No breast-feeding	53.6	32.6	13.6	6.5	18.2	4.4	15.2	6.5	4.6	15.2
Any breast-feeding	42.7*	31.5	14.5	7.3	13.6	10.8	5.7***	3.0	7.1	8.2**
1–5 mo	42.5	28.8	11.9	8.5	11.9	9.3	9.8	4.6	7.7	4.6
6–11 mo	42.9	30.4	14.6	7.4	13.7	12.0	5.5	2.9	7.8	5.7
12+ mo	42.8	34.8	16.2	6.3	14.6	10.7	3.0****	1.9*	6.1	13.8

Note. Data were missing on cause of death for 20 males and 14 females. Injury = accident, suicide, or homicide. *P* values are for the breast-feeding category shown compared with no breast-feeding, based on chi-square test (any breast-feeding) or Mantel-Haenzsel chi-square test for trend (duration).
*P = .08; **$P \leq .11$; ***P = .01; ****$P \leq .001$.

TABLE 4—Age-Adjusted and Multiply-Adjusted Relative Hazards (RHs) for Breast-Feeding Duration and Total and Cause-Specific Mortality, by Sex: Subjects in the Terman Life Cycle Study

Breast-Feeding Duration	All Causes		Cardiovascular Disease		Cancer		Injury		Other	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
Age-adjusted RH										
Any breast-feeding	0.75*	1.02	0.98	1.18	0.68	2.57	0.37***	0.47	1.45	0.56*
1–5 mo	0.80	0.94	0.88	1.43	0.66	2.19	0.64	0.76	1.65	0.32**
6–11 mo	0.75*	0.97	0.97	1.17	0.67	2.77	0.35***	0.46	1.55	0.38**
12+ mo	0.74*	1.13	1.07	0.99	0.67	2.67	0.20****	0.28*	1.21	0.96
Multiply-adjusted RH[a]										
Any breast-feeding	0.77*	1.20	1.04	1.07	0.64*	2.25	0.43**	1.00	1.36	0.73
1–5 mo	0.81	1.10	0.92	1.28	0.65	1.89	0.67	1.54	1.50	0.43
6–11 mo	0.77	1.16	1.05	1.06	0.62	2.32	0.40**	1.01	1.58	0.56
12+ mo	0.74*	1.36	1.13	0.87	0.64	2.57	0.25***	0.58	1.00	1.25

Note. Data were missing on cause of death for 20 males and 13 females and on the control variables for 25 males and 25 females. *P* values are for the breast-feeding category shown compared with no breast-feeding, based on Cox proportional hazards model.
[a]Adjusted for age at baseline, birthweight, infant health, and childhood socioeconomic status.
*P = .11–.15; **$P \leq .05$; ***$P \leq .01$; ****$P \leq .001$.

with and without control for age, birthweight, infant health, and childhood socioeconomic status. No adjustments were made for multiple comparisons; rather detailed *P* values are presented. All *P* values are for two-tailed tests.

There were 276 siblings, representing 105 families, within the sample. Analyses restricted to only one sibling per family yielded similar results; therefore, only analyses of the entire sample are presented.

Results

Of the 1170 subjects, 69 boys (10.7%) and 46 girls (8.8%) had never been breast-fed. As can be seen in Table 1, the distribution of total duration of breast-feeding did not differ significantly by sex. The total number of months subjects were breast-fed ranged up to 24 months for boys (mean = 8.2) and up to 36 months for girls (mean = 8.0).

The distribution of potentially confounding variables of the breast-feeding and longevity analysis is presented in Table 2 by breast-feeding duration. A trend analysis is significant for age for males and for birthweight and socioeconomic status for females. Of these variables, only birthweight demonstrates a clear gradient with breast-feeding duration. The significant association for age reflects sample size more than a strong association, and the significance for socioeconomic status reflects the difference in one category more than a trend across breast-feeding categories. Table 2 also demonstrates that among infants who survived to childhood or adolescence, breast-feeding was associated with better infant health among both boys and girls, with a clear gradient of better health with increasing duration of breast-feeding. After adjustment for differences in age, birthweight, and childhood socioeconomic status, variables that might influence both breast-feeding practices and health, the association with better infant health remained significant for boys and nearly reached significance for girls. The odds ratio of having poor or fair infant health for those never breast-fed compared with those breast-fed for 12 or more months was 4.00 ($P < .001$) for boys and 2.38 ($P \approx .10$) for girls. These odds may reflect a causal association or the possibility that sick infants were less likely to be breast-fed. An analysis of breast-feeding

duration and each of the six personality traits (not shown) demonstrated no significant associations.

By 1986, 284 of the men (43.9%) and 165 of the women (31.5%) had died. Total mortality rates were lower among men, but not women, who had ever been breast-fed compared with those who had not, but they did not vary systematically with breast-feeding duration for either sex (Table 3). Of the men, 90 had died from cardiovascular disease (55 of these from coronary heart disease), 88 from cancer, 42 from injury, and 43 from other causes. Of the women, 37 had died from cardiovascular disease (15 from coronary heart disease), 52 from cancer, 17 from injury, and 45 from other causes. Cause of death was unknown for 21 men and 14 women. Cause-specific mortality rates demonstrated that only the association between duration of breast-feeding and deaths from injuries was consistent between men and women.

Table 4 presents the survival analysis over the average 65 years of follow-up for breast-feeding and breast-feeding duration, adjusted for differences in age at baseline as well as birthweight, infant health, and childhood socioeconomic status. Although longer overall survival was associated with any breast-feeding for men in this analysis (relative hazard = .75), the difference was not statistically significant when adjusted for age or other covariates ($P = .15$) and there was only a weak association with duration. The opposite association between breast-feeding and survival for women did not approach statistical significance.

Breast-feeding was not associated with survival from cardiovascular disease for either sex. This was true both before and after adjustment for differences in age at baseline, birthweight, infant health, and childhood socioeconomic status. Similarly, no association was seen for coronary heart disease (data not shown). Although breast-feeding was associated with increased survival from cancer for men and decreased survival for women, the association was not significant for either sex, and there was no significant association with duration either before or after adjustment for potentially confounding variables. Breast-feeding was associated with increased survival from injury for men ($P = .03$), with a clear and significant gradient of longer survival with increasing duration of breast-feeding. Women who had been breast-fed for 12 months or longer also had longer survival ($P = .11$), although this association did not approach statistical significance after full adjustment.

Discussion

In the present study, we found a weak association between breast-feeding and increased longevity among men before and after adjustment for age, birthweight, infant health, and childhood socioeconomic status. Likewise, there was only a weak association with duration of breast-feeding. No association was found among women. Neither cardiovascular disease nor cancer survival demonstrated a significant association with duration of breast-feeding for either sex. Survival from deaths due to injury, however, was positively associated with breast-feeding among men, before and after adjustment for potentially confounding variables, and demonstrated a significant gradient with duration of breast-feeding.

The lack of an association between breast-feeding and subsequent cardiovascular or coronary heart disease survival is contrary to the findings of Barker and colleagues.[7–9] In their study, 5654 men born in England between 1911 and 1930 were traced for mortality between 1951 and 1987 (ages 20 to 74 years). Standardized mortality ratios indicated that men who had been breast-fed had lower ischemic heart disease mortality rates than men who had been bottle-fed (standardized mortality ratios of 81 and 94, respectively).[7] Subsequent analyses in a subset of 485 men indicated that the benefit occurred only among men who had been both breast- and bottle-fed or breast-fed and weaned before their first birthday (standardized mortality ratios of 73 and 79, respectively) compared with men who had been breast-fed and not weaned (standardized mortality ratios = 97) or bottle-fed exclusively (standardized mortality ratios = 95).[9] The results were not adjusted for differences in socioeconomic status. Differences between these two studies may reflect cultural differences between breast-feeding in England and the United States in the early 1900s (i.e., differences in supplementation with bottles, the timing of addition of solid food, or the association of socioeconomic status with breastfeeding practices). Alternatively, the differences may be due to sample sizes or chance.

In the subset of men born in England,[9] those who had been breast-fed and weaned before their first birthday and those who had been both breast- and bottle-fed (i.e., those who had the lowest risk of death from ischemic heart disease) had lower serum total cholesterol, LDL levels, and ratios of LDL to HDL as adults than did men who had been breast-fed and not weaned before their first birthday or who had been exclusively bottle-fed. A study of 30-year-olds found that women who had been breast-fed as infants had lower levels of total cholesterol than women who had been bottle-fed, by an average of 20 mg/dL. The difference among men, however, was only 2 mg/dL.[17] Differences in serum cholesterol levels are small and inconsistent when comparing children and young adults who were breast-fed or bottle-fed as infants.[10] A recent study of 3- to 4-year-old children found no association between breast-feeding duration and diastolic or systolic blood pressure and only a weak positive association with resting heart rate.[18]

There has been no previously published finding of an association between breast-feeding and lower risk of death from injury. The lack of a strong association among women may indicate that the association among men is due to chance; however, women who had been breast-fed for 12 or more months also demonstrated a reduction in risk, although it was not significant. Perhaps mothers who breast-fed in the early 1900s developed stronger psychosocial ties with their children (especially sons), which in turn better prepared the children to interact with others and thereby avoid both suicide and homicide. Perhaps mothers who breast-fed their infants were more safety conscious and therefore raised children who were less prone to accidents. Or perhaps some children (that is, those who were more docile) were both more likely to breast-feed longer and less likely to take chances as they got older. However, in the present study we found no association between being breast-fed and conscientiousness or any other personality trait measured in childhood.

The present sample is restricted to intellectually gifted children from predominantly middle-class families, who tended to become high achievers.[13] Therefore, these findings may not be generalizable to the entire population. However, because of the nature of this sample, several potentially confounding influences can be ruled out. For example, any association between breast-feeding and longevity is probably not confounded by access to medical care, given that the sample was largely middle-class. Recall

bias is also unlikely, because the breast-feeding information was provided by the parent, on average within 10 years of birth and long before the subject's death. It is unlikely that infant health was a confounding factor, for several reasons: all analyses were adjusted for infant health status; only children who survived infancy and early childhood were included in the sample; and in the early 1900s, children too sick to nurse were less likely to survive and to be subsequently identified as intellectually gifted.

The present analysis focused on whether breast-feeding in the early 1900s was associated with adult longevity. Only children who survived infancy and early childhood were included in the sample; therefore, associations do not reflect known associations with early mortality but only those with adult longevity. Although breast-feeding practices have changed over time, any biologically mediated association should still be evident.

Overall, the present study does not provide strong evidence that breast-feeding is related to adult longevity. The association of breast-feeding with deaths due to injury among men may be due to chance alone or to psychosocial correlates of breast-feeding practices. □

Acknowledgments

This research was supported by National Institute on Aging grant AGO8825.

Some of the data were made available from the Terman Life Cycle Study, begun by Lewis Terman in 1921. Further assistance was provided by Eleanor Walker of the Terman Project.

References

1. Report of the Task Force on the Assessment of the Scientific Evidence Relating to Infant-Feeding Practices and Infant Health. *Pediatrics.* 1984;74:579–762.
2. Cunningham AS, Jelliffe DB, Jelliffe EFP. Breastfeeding and health in the 1980s: a global epidemiologic review. *J Pediatrics.* 1991;118:659–666.
3. Davis MK, Savitz DA, Graubard BI. Infant feeding and childhood cancer. *Lancet.* 1988;2:365–368.
4. Mayer EJ, Hamman RF, Gay EC, et al. Reduced risk of IDDM among breastfed children. *Diabetes.* 1988;37:1625–1632.
5. Koletzko S, Sherman P, Corey M, et al. Role of infant feeding practices in development of Crohn's disease in childhood. *BMJ.* 1989;298:1617–1618.
6. Whorwell PJ, Holdstock G, Whorwell GM, et al. Bottle feeding, early gastroenteritis, and inflammatory bowel disease. *BMJ.* 1979;1:382.
7. Barker DJP, Winter PD, Osmond C, et al. Weight in infancy and death from ischaemic heart disease. *Lancet.* 1989;2:577–580.
8. Barker DJP. The intrauterine environment and adult cardiovascular disease. In: Bock GR, Whelan J, eds. *The Childhood Environment and Adult Disease.* Chichester, England: John Wiley & Sons Inc; 1991:3–16. Ciba Foundation Symposium No. 156.
9. Fall CHD, Barker DJP, Osmond C, et al. Relation of infant feeding to adult serum cholesterol concentration and death from ischaemic heart disease. *BMJ.* 1992;304:801–805.
10. Hamosh M. Does infant nutrition affect adiposity and cholesterol levels in the adult? A review. *J Pediatr Gastroenterol Nutr.* 1988;7:10–16.
11. Mott GE, Lewis DS, McGill HC Jr. Programming of cholesterol metabolism by breast or formula feeding. In: Bock GR, Whelan J, eds. *The Childhood Environment and Adult Disease.* Chichester, England: John Wiley & Sons Inc; 1991:56–76. Ciba Foundation Symposium No. 156.
12. Terman LM, Oden MH. *Genetic Studies of Genius, IV: The Gifted Child Grows Up: Twenty-Five Years Follow-Up of a Superior Group.* Stanford, Calif: Stanford University Press; 1947.
13. Terman LM, Oden M. *Genetic Studies of Genius, V: The Gifted Group at Mid-Life.* Stanford, Calif: Stanford University Press; 1959.
14. Friedman HS, Tucker J, Tomlinson-Keasey C, et al. Does childhood personality predict longevity? *J Pers Soc Psychol.* 1993;65:176–185.
15. *The International Classification of Diseases,* Ninth Revision, Clinical Modification. 2nd ed. Hyattsville, Md: National Center for Health Statistics; 1980. DHHS publication PHS 80-1260.
16. Cox DR. Regression models and life tables (with discussion). *J R Stat Soc B.* 1972;8:187–220.
17. Marmot MG, Page CM, Atkins E, Douglas JBW. Effect of breast feeding on plasma cholesterol and weight in young adults. *J Epidemiol Community Health.* 1980;34:164–167.
18. Baranowski T, Bryan GT, Harrison JA, et al. Height, infant-feeding practices and cardiovascular functioning among 3 or 4 year old children in three ethnic groups. *J Clin Epidemiol.* 1992;45:513–518.

Harsh Social Conditions and Perinatal Survival: An Age-Period-Cohort Analysis of the World War II Occupation of Norway

Allen J. Wilcox, MD, PhD, Rolv Skjærven, PhD, and Lorentz M. Irgens, MD, PhD

ABSTRACT

Objectives. The hypothesis was tested that unfavorable social conditions are associated with poor perinatal survival through direct effects on pregnancy or, more indirectly, through effects on mothers born under such conditions. The occupation of Norway by Nazi Germany was used as a period of social hardship.

Methods. Data from Norwegian vital statistics and the Medical Birth Registry were used to describe perinatal mortality during World War II and also a generation later, among babies born to mothers who had themselves been born during the war. Logistic regression was used to identify a possible cohort effect among mothers born in 1940 through 1944 compared with mothers born before or after that period.

Results. Harsh conditions in Norway during the occupation increased childhood mortality. However, perinatal mortality declined during that period. Likewise, no adverse effect was seen on the survival of babies born to mothers who had themselves been born during the war (odds ratio = 1.00; 95% confidence interval = 0.96, 1.04).

Conclusions. We find no evidence that wartime conditions in Norway impaired perinatal survival, either directly or through an effect on women born during the war. These data underscore how little is known about the ways that social conditions influence perinatal mortality. (*Am J Public Health.* 1994;84:1463–1467)

Introduction

In this century, rates of perinatal mortality have fallen steeply in Norway, as they have in all developed countries. The precise causes of this secular decline are unclear; however, there is general agreement that broad improvements in living conditions, nutrition, and public health have contributed.[1] A similar decline in perinatal mortality took place a century earlier among the European aristocracy, before the era of modern medical care.[2] This supports the hypothesis that social factors make sizable contributions to perinatal survival.[3]

Social conditions may have immediate or delayed influence on perinatal survival. Some factors may act directly, around the time a woman becomes pregnant and gives birth (a period effect). Severe malnutrition is a well-documented example of such an effect.[4,5] Other factors (perhaps including nutrition) may accumulate over the lifetime of the mother (a cohort effect). As a specific type of cohort effect, some authors have hypothesized that a woman's earliest experience as a fetus and young child contributes to her ultimate reproductive success.[6–8] Although all these hypotheses have some plausibility, the usual vital-statistics data offer little basis for preferring one interpretation over another. Without conclusive evidence, the common assumption is that an infant's survival is most strongly affected by conditions prevailing at the time of its fetal development and birth.[9,10]

An historical event may shed light on this issue. In 1940, Norway was invaded and occupied by Nazi Germany. For the next 5 years, the Norwegian population was subjected to severe food shortages and other disruptions. We looked for immediate and delayed effects of those harsh social conditions on perinatal health.

Methods

We drew on two sources of Norwegian perinatal data. One is the published tables of vital statistics, which include annual rates of infant mortality since 1836.[11] We present perinatal and childhood mortality rates for a 50-year period including World War II (1921–1970). Those data are straightforward and require no special analysis.

The second source of data is the Medical Birth Registry of Norway. The Medical Birth Registry comprises all births (including fetal deaths with gestational ages of 16 weeks or more) recorded in Norway since 1967. The annual number of births in the registry has ranged from 50 000 (in 1983) to 69 000 (in 1969), totaling 1.45 million births through 1991. All infant deaths are routinely linked to the birth registry by using the national identification number. The registry includes mother's birth date, from which her birth cohort and age at delivery can be determined.

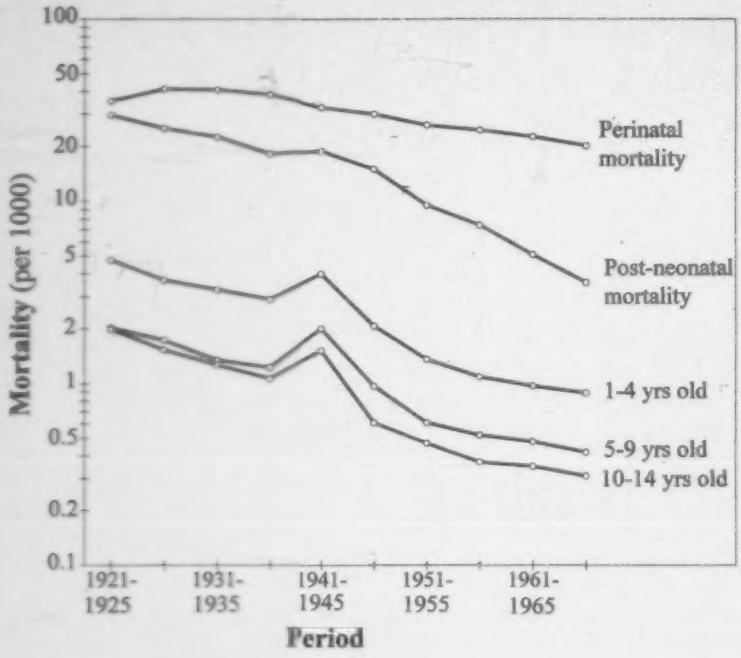
Mothers with babies in the Medical Birth Registry were themselves born over a range of years that spans World War II. To carry out an analysis of births among

Allen J. Wilcox is with the Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC. Rolv Skjærven is with the Section for Medical Informatics and Statistics and the Medical Birth Registry, University of Bergen, Bergen, Norway. Lorentz M. Irgens is with the Medical Birth Registry and the Division of Preventive Medicine, Department of Public Health and Primary Health Care, University of Bergen.

Requests for reprints should be sent to Allen J. Wilcox, MD, PhD, Epidemiology Branch, NIEHS, Research Triangle Park, NC 27709.

This paper was accepted May 17, 1994.

Editor's Note. See related editorial by Paneth (p 1372) in this issue.



Source. Based on data from Official Statistics of Norway.[11]

FIGURE 1—Perinatal, postneonatal, and childhood mortality, Norway, 1921 through 1970.

the cohort of mothers born during the war, births in the Registry were categorized into 5-year maternal birth cohorts (<1925, 1925–1929, 1930–1934 . . . 1965–1969, 1970–1974, and 1975+), as well as by the year of the baby's birth, also in 5-year time periods (1967–1971, 1972–1976, 1977–1981, 1982–1986, and 1987–1991). Five-year categories for birth periods and maternal birth cohorts define maternal-age groups that correspond approximately—but not precisely—to 5-year categories. This feature of cohort analysis has been clearly described by Case.[12] For example, births in 1967 through 1971 to women born in 1940 through 1944 will be chiefly to women aged 25 through 29 years, but will include some women as young as 23 or as old as 31. Although it is not always discussed explicitly in age-period-cohort analyses, imprecision is inherent in the grouping process. We note this imprecision by designating each maternal-age category according to its central year, for example as "age 27" rather than as "ages 25 through 29."

In the cohort analysis, we first looked separately at stillbirths and at early neonatal deaths (deaths in the first week of life). Results were similar for the two outcomes, and so these two categories of death were combined as perinatal mortality. Rates for mothers younger than 15 or older than 44 years are not included because of instability due to small numbers.

It is a property of age-period-cohort data that when any two of the three variables are fixed, the third is determined. Thus, period and cohort effects are inherently dependent on each other, and no simple solutions are available to separate them.[13] Determination of a possible effect through a specific cohort can be approached as follows. First, we used a logistic regression analysis to model the general effects of birth period and maternal age (as categorical covariates) on perinatal mortality.[14] This simple model provided a reasonable fit to the whole data set ($\chi^2 = 22.8$, 20 *df*, $P = .30$). A single linear (or "drift") term for the period effect reduced the goodness of fit considerably ($\chi^2 = 86.5$, 34 *df*, $P < .01$).[15]

To test for a specific cohort effect in women born during the war, a "dummy" variable was created comprising the five period-age combinations that correspond to the war cohort. This dummy variable includes one data point from each relevant period or, equivalently, one data point from each relevant maternal-age group. The overall effects of age and period were then reestimated for the whole data set excluding the dummy cohort variable. Finally, the possible influence of the war cohort variable (in addition to the effects of age and period) was assessed by likelihood ratio tests.[14] Seen another way, this process estimated a cohort time trend based on all cohorts born before or after World War II, and then measured any deviance of the war cohort from the expected trend. For completeness, the same test was performed separately on each birth cohort.

Results

Mortality rates for Norwegian infants and children from 1921 through 1970 are shown in Figure 1. Mortality is on a log scale to accommodate the wide range of rates. Death rates declined in all categories by at least 50% during the time period. The most striking effects of World War II were among children over 1 year of age. Within each age category of childhood mortality, the death rate increased by at least a third and then fell in the next time period to below prewar levels. These patterns were consistent for both boys and girls and for deaths due to accident as well as illness (data not shown). For postneonatal mortality (deaths between 1 month and 1 year of age), the overall decline was interrupted only slightly during the war years. There is no evidence for any effect of World War II on the decline of perinatal mortality (stillbirths plus deaths in the first week of life).

Table 1 presents data from the Medical Birth Registry on births among mothers who were born from 1920 through 1974, the same general time period as Figure 1. Data are sorted by age of the mother, birth cohort of the mother, and period of the birth. Columns correspond to maternal age, horizontal rows show the mother's birth cohort, and diagonals (lower left to upper right) show the time period during which the index birth occurred. Table 2 presents perinatal death rates (per 1000) within the same groups. Declining mortality rates are apparent with advancing period and cohort. Risk is elevated among the youngest and oldest mothers.

TABLE 1—Distribution of Live Births Plus Stillbirths, by Maternal Age, Cohort of Maternal Birth, and Period of Index Birth, Norway, 1967 through 1991

Period of Index Birth	Maternal Age, y						Maternal Birth Cohort
	17	22	27	32	37	42	
						9 770	1925–1929
					25 781	4 724	1930–1934
				51 254	16 670	2 977	1935–1939
			98 423	47 155	15 508	3 390	1940–1944
		120 478	106 019	52 697	20 032	4 726	1945–1949
	29 602	97 420	90 432	56 831	26 019		1950–1954
	26 191	78 933	95 039	75 251			1955–1959
1967–1971	17 717	67 487	106 161				1960–1964
1972–1976	13 971	70 006					1965–1969
1977–1981	12 410						1970–1974
1982–1986							
1987–1991							

Note: Horizontal rows show mother's birth cohort, and diagonal rows (lower left in the first column of the table to the last column under "Maternal Age") show the period during which the index birth occurred.

TABLE 2—Perinatal Mortality per 1000 Live Births and Stillbirths within Categories of Maternal Age, Cohort of Maternal Birth, and Period of Index Birth, Norway, 1967 through 1991

Period of Index Birth	Maternal Age, y						Maternal Birth Cohort
	17	22	27	32	37	42	
						43.2	1925–1929
					30.7	36.2	1930–1934
				23.4	28.7	27.5	1935–1939
			20.6	20.3	21.2	20.6	1940–1944
		21.5	17.9	15.1	17.3	24.8	1945–1949
	25.2	17.9	14.4	14.3	17.8		1950–1954
	22.8	14.6	11.7	13.4			1955–1959
1967–1971	16.9	12.8	11.4				1960–1964
1972–1976	17.8	11.7					1965–1969
1977–1981	14.7						1970–1974
1982–1986							
1987–1991							

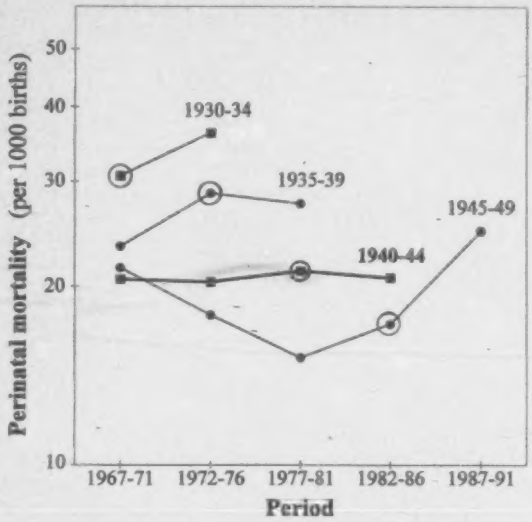
Note: Horizontal rows show mother's birth cohort, and diagonal rows (lower left in the first column of the table to the last column under "Maternal Age") show the period during which the index birth occurred.

The decline in perinatal mortality seen in Figure 1 has continued throughout the subsequent 25 years covered by the registry. Figure 2 shows perinatal mortality among four birth cohorts of women, arranged by the time period in which their pregnancies occurred. Two cohorts were born before World War II, one was born during the war, and the fourth was born immediately after. There is a general pattern of decreasing mortality within each successive cohort, although this pattern is complicated by maternal age. In any given year, women from different birth cohorts also differ in age. As an illustration of this, the perinatal mortality for mothers in the 37-year age group is circled on each cohort curve. Figure 3 shows the cohort curves for all birth cohorts arranged by mother's age.

Our hypothesis was that the babies of women born in 1940 through 1944 would have higher perinatal mortality than that predicted from the cohorts of mothers born before and after the war. We used an analytic approach that tested whether the experience of the war cohort differed from the overall pattern of steadily improving mortality in successive cohorts. After modeling the general effects of age and period, the odds ratio for the war cohort was 1.00 (95% confidence interval = 0.96, 1.04). Thus, we found no evidence that being born during World War II impaired women's ability to bear healthy infants.

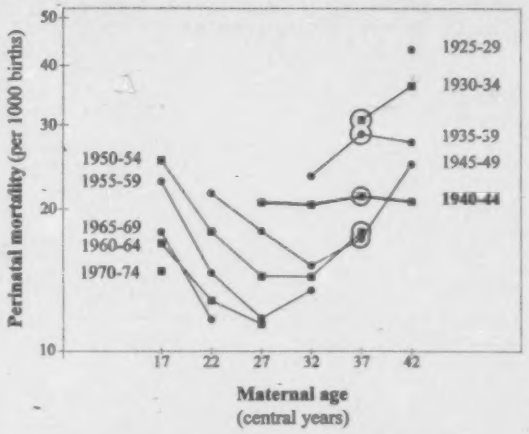
We carried out the same analytic procedure on the other birth cohorts represented in the registry. Odds ratios for these cohorts ranged from 0.99 to 1.02, with tight confidence intervals due to the large number of births within each cohort. The reproductive success of each cohort of mothers is what would be predicted by the observed mortality trends associated with period and age.

Of course, this analytic approach cannot distinguish the relative contributions of period and cohort to perinatal mortality; it shows only that when taken one at a time, no single maternal cohort deviates from the pattern set by all other cohorts in the sample. Another limitation of this test is that it does not consider



Note. Circles indicate mortality for mothers with a central age of 37 years.
Source. Based on data from the Medical Birth Registry of Norway.

FIGURE 2—Cohort-specific perinatal mortality for selected maternal cohorts, by period, Norway, 1967 through 1991.



Source. Based on data from the Medical Birth Registry of Norway.

FIGURE 3—Cohort-specific perinatal mortality, by maternal age, Norway, 1967 through 1991.

ways in which mortality patterns *within* cohorts might differ. For example, the war cohort (1940–1944) has a relatively flat pattern of perinatal mortality with advancing age compared with other cohorts. The meaning of this pattern, if any, is unclear.

Discussion

Perinatal mortality is continuing to decline in many countries, which is an expression of improvements in social and other factors that determine perinatal risk. Within this pattern of general improvement, large differences in perinatal mortality by socioeconomic status persist. How important are social conditions around the time of pregnancy or conditions earlier in the life of the mother? Norwegian data allow both factors to be studied in relation to the hardship of World War II. During World War II, the Norwegian population suffered not only deprivation of food and other essentials, but also punishment for resistance to German occupation. Many were imprisoned, and war-related deaths are estimated at about 10 000.

We looked for wartime effects on perinatal mortality during that time and also a generation later, when women born during the occupation themselves gave birth. Data for the latter group of births are not routinely available; vital statistics do not link the mother's age with survival of her infant, and so perinatal outcomes for maternal birth cohorts must be constructed from other sources. The Medical Birth Registry of Norway provides the mother's age linked with perinatal mortality for a recent 25-year period and includes many mothers born during World War II.

We were able to find no evidence that Norway's experience during World War II affected perinatal survival, either immediately or through indirect effects on mothers born during the war. These data do not contradict previous reports of increased perinatal death associated with maternal starvation. Maternal malnutrition appears to impair perinatal survival when caloric intake falls below a certain threshold,[5] and extreme malnutrition was rare in Norway. However, the disruption and distress that accompanied German occupation are unquestioned, and the absence of measurable effects on perinatal mortality raises several possible interpretations.

1. Deaths may have been incompletely registered during this period, especially fetal or early infant deaths. By 1940, the registry of perinatal deaths had been routine in Norway for a century. We know of no evidence to suggest that registration deteriorated during the war; still, that possibility cannot be ruled out. Underregistration of deaths might have contributed to the absence of a direct war effect on perinatal mortality; it would not account for the absence of an effect a generation later.

2. Pregnant women and infants may have been relatively protected from the hardships of the occupation. It is difficult to reconstruct all aspects of this disrupted period of history, now a half-century in the past. Norway was not self-sufficient in food production; during the blockade of imports in 1942 through 1945, severe undernutrition and frank malnutrition were commonplace.[16] Pregnant women and infants probably did not suffer the worst; however, deprivation was widespread. Young children were not spared, as suggested by their mortality rates (Figure 1). Childhood growth retardation and delay of menarche during this time have been documented.[17,18] Because conditions were most severe in the last part of the war, we carried out additional analyses in 3-year groupings, separating the 1940 through 1942 cohort from the 1943 through 1945 cohort. Results were unchanged; no cohort effect was apparent in either group.

3. To produce a cohort effect on reproduction, hardship may have to come

later in childhood, or be sustained over a longer time. Logistic regression analysis of data from cohorts of women who were born before the war (and thus were older children at the time of the war) showed no evidence of adverse effects on their later reproduction. Possible effects of more prolonged deprivation cannot be evaluated with these data.

4. War conditions may have discouraged some high-risk women from becoming pregnant, thus concealing adverse wartime effects on those pregnancies that actually occurred. Maternal age at delivery did not change in Norway between 1935 through 1939 and 1940 through 1944.[9] Furthermore, Norwegian age-specific fertility rates were actually higher during the war than in the previous 5-year period.[9] Thus, we find no gross evidence of self-selection for pregnancy related to the war. The influence of more subtle selection among women who became pregnant is difficult to assess.

5. Perinatal mortality may be too crude an end point; period or cohort effects may be detectable only among less drastic outcomes such as fertility or fetal growth. Fertility or fetal growth effects may occur[19] but are not directly relevant to the present analysis. The object here was to explore connections between social conditions and perinatal survival.

Some additional data are available from Holland, which also had severe wartime experience. Parts of Holland suffered extreme deprivation during the German occupation. Famine occurred in some regions between October 1944 and May 1945. Pregnant women who were starved in their first trimester experienced increased rates of stillbirth, whereas those who were starved late in pregnancy had increased rates of infant death in the first 3 months of life.[5] Subsequent cohort effects on perinatal survival of babies born to these Dutch women have been suggested but are not well documented.[20] Unlike the experience in Norway, national rates of postneonatal mortality in Holland jumped sharply during the war to more than double the prewar level.[21] However, national rates of perinatal mortality declined for Holland, as they did in Norway.[21]

Maternal age, birth period, and maternal birth cohort are not isolated variables, of course. For instance, maternal age is correlated with parity, and parity is in turn affected by desired family size and selective fertility,[22] which change with time. These and other variables may be relevant to the interpretation of period and cohort effects, but they are beyond the scope of the present analysis.

In conclusion, we looked for increases of perinatal mortality specifically related to the occupation of Norway by Nazi Germany during World War II. Whereas mortality increased for children during the war, there was only a slight effect on postneonatal deaths and none on perinatal deaths. We found no evidence that women who were born during the war suffered later consequences in their ability to bear viable infants. Our inability to detect wartime effects underscores how little we understand the ways in which social factors act on perinatal mortality. □

Acknowledgments

We are grateful to Donna Baird, Tine Brink Henriksen, Irva Hertz-Picciotto, Rolv Terje Lie, Ruth Little, L. H. Lumey, Andrew Rowland, David Umbach, Clarice Weinberg, and Pat Schreuder, who made valued contributions to this paper.

References

1. Yankauer A. Infant mortality and morbidity in the International Year of the Child. *Am J Public Health.* 1979;69:852–853.
2. Peller S. Studies on mortality since the Renaissance. *Bull Hist Med.* 1943;13:427–461.
3. Wilcox A. Infant mortality revisited. *Perinat Epidemiol.* 1993;7:347–348.
4. McCance RA, Widdowson EM. The determinants of growth and form. *Proc R Soc Lond, B.* 1974;185:1–17.
5. Stein ZA, Susser M, Saenger G, et al. *Famine and Human Development: The Dutch Hunger Winter of 1944–45.* New York: Oxford University Press; 1975.
6. Emanuel I, Hale CB, Berg CJ. Poor birth outcomes of American Black women: an alternative explanation. *J Public Health Policy.* 1989;10:299–308.
7. Emanuel I, Filakti H, Alberman E. Intergenerational studies of human birthweight from the 1958 cohort, I: evidence for a multigenerational effect. *Br J Obstet Gynaecol.* 1992;99:67–74.
8. Baird D. Epidemiologic patterns over time. In: Reed DM, Stanley FJ, eds. *The Epidemiology of Prematurity.* Baltimore, Md: Urban & Schwarzenberg; 1977:5–14.
9. National Center for Health Statistics. Infant mortality problems in Norway. *Vital Health Stat [3].* 1967:8.
10. Hoffman HJ, Meirik O, Bakketeig LS. Methodological considerations in the analysis of perinatal mortality rates. In: Bracken M, ed. *Perinatal Epidemiology.* New York: Oxford University Press; 1984:491–530.
11. *Historical Statistics 1978.* Oslo, Norway: Central Bureau of Statistics of Norway; 1978.
12. Case RAM. Cohort analysis of mortality rates as an historical or narrative technique. *Br J Prev Soc Med.* 1956;10:159–171.
13. Holford TR. Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annu Rev Public Health.* 1991;12:425–457.
14. Dixon WJ, chief ed. *BMDP Statistical Software Manual.* Los Angeles, Calif: University of California Press; 1992.
15. Clayton D, Schifflers E. Models for temporal variation in cancer rates, I: age-period and age-cohort models. *Stat Med.* 1987;6:449–467.
16. Gogstad AC. *Helse og Hakekors.* Bergen, Norway: Alma Mater; 1991.
17. Brundtland GH, Liestøl K, Walløe L. Height, weight and menarcheal age of Oslo school children during the last 60 years. *Ann Hum Biol.* 1980;7:307–322.
18. Bjelke E. Variation in height and weight in the Norwegian population. *Br J Prev Soc Med.* 1971;25:192–202.
19. Lumey LH. Decreased birthweights in infants after maternal *in utero* exposure to the Dutch famine of 1944–1945. *Pædiatr Perinat Epidemiol.* 1992;6:145–152.
20. Kloosterman GJ. Social factors and their influence on perinatal mortality. In: Proceedings of the 8th World Congress of Gynæcology and Obstetrics; October 17–23, 1976, Mexico City, Mexico.
21. National Center for Health Statistics. International comparison of perinatal and infant mortality: the United States and six West European Countries. *Vital Health Stat [3].* 1967;6.
22. Skjærven R, Wilcox AJ, Lie RT, Irgens LM. Selective fertility and the distortion of perinatal mortality. *Am J Epidemiol.* 1988;128:1352–1363.

State Family Planning and Abortion Expenditures: Their Effect on Public Health

Kenneth J. Meier, PhD, and Deborah R. McFarlane, DrPH, MPA

ABSTRACT

Objectives. This study examines whether state family planning expenditures and abortion funding for Medicaid-eligible women might reduce the number of low-birthweight babies, babies with late or no prenatal care, and premature births, as well as the rates of infant and neonatal mortality.

Methods. Using a pooled time-series analysis from 1982 to 1988 with the 50 states as units of analysis, this study assessed the impact of family planning expenditures and abortion funding on several public health outcomes while controlling for other important variables and statistical problems inherent in pooled time-series studies.

Results. States that funded abortions had a significantly higher rate of abortions and significantly lower rates of teen pregnancy, low-birthweight babies, premature births, and births with late or no prenatal care. States that had higher expenditures for family planning had significantly fewer abortions, low-birthweight babies, births with late or no prenatal care, infant deaths, and neonatal deaths.

Conclusions. Funding abortions for Medicaid-eligible women and increasing the level of expenditures for family planning are associated with major differences in infant and maternal health in the United States. (*Am J Public Health.* 1994;84:1468–1472)

Introduction

Except for the obvious impact on maternal mortality,[1] the debate over government funding for abortions has seldom focused on public health. Instead, the arguments usually pit the equal rights of poor women to control their own bodies against the rights of those taxpayers opposed to abortion. Currently, the Hyde amendment prohibits the federal government from funding abortions except in cases of incest or rape.[2] (Recent proposed regulations would require states to fund abortions under these conditions.) States, however, may use their own funds to pay for abortions for Medicaid-eligible women, and 12 states currently do so.

Previous studies suggest that the availability of both abortion and family planning services produces public health benefits. Grossman and Jacobowitz found that increased access to legal abortion was the most important factor in the large decreases in US neonatal mortality from 1964 to 1977[3]; the second most important factor was subsidized family planning services for poor women. In a subsequent study, Corman and Grossman reported that increased availability of abortion was by far the most important factor in the decline in Black neonatal mortality[4]; for Whites, however, abortion availability was only the fifth leading contributor to lower neonatal rates. Abortion affects neonatal mortality by reducing the number of unwanted births. Because unwanted births are less likely to receive early prenatal care and may be linked to low birthweights,[5–8] access to abortion should be indirectly associated with earlier prenatal care and fewer low birthweights. Both reduce neonatal mortality. (It has been found that Medicaid restrictions on abortion funding increase the probability of birth but have no impact on birthweights.[8])

A 1984 study found that states that did not have public funds available to pay for abortions for low-income women had to spend more money "to provide maternity care, medical care for the infant, Aid to Families with Dependent Children, and nutritional assistance to women on Medicaid."[9(p111)] For every dollar used to pay for abortions for poor women, the study found that more than $4 were saved in medical and social welfare costs over the next 2 years.[9]

The primary purpose of this paper is to examine whether state funding for abortions and family planning from 1982 to 1988[10] actually produced public health benefits. Both policies attempt to reduce unwanted pregnancies so that a larger percentage of the pregnancies that are carried to term are wanted. This should reduce prenatal problems and result in healthier babies.

Data Sources and Methods

Dependent Variables: Public Health Impacts

State abortion rates—the number of abortions per 1000 women aged 15 to 44—were obtained from the Alan Guttmacher Institute.[11,12] To provide estimates for 1983 and 1986, when the institute did not collect data, the McFarlane and

Kenneth J. Meier is with the Department of Political Science at the University of Wisconsin–Milwaukee. Deborah R. McFarlane is with the University of New Mexico, Albuquerque.

Requests for reprints should be sent to Kenneth J. Meier, PhD, Department of Political Science, University of Wisconsin–Milwaukee, Box 413, Milwaukee, WI 53201.

This paper was accepted May 5, 1994.

Editor's Note. See related editorial by Klerman and Klerman (p 1377) in this issue.

Meier procedure was used.[13] For 1983 data, the ratio of abortions reported to the Centers for Disease Control (CDC) in 1983 to those reported in 1982 was calculated for each state. This procedure assumes that CDC abortion data, while biased, are biased consistently from year to year. This ratio, the state's annual growth rate in abortions, was then multiplied by 1982 abortion data from the institute to estimate the number of abortions in each state for 1983. These figures were then converted into rates by dividing them by the number of women aged 15 to 44 (in thousands) in that state. Similar estimates for 1986 were constructed using 1985 and 1986 CDC data.

All other dependent variables were taken from the National Center for Health Statistics for the years 1982 through 1988. These measures are the teen birthrate (number of births to women under age 20 per 1000 women aged 14 to 19), the percentage of low-birthweight babies (<2.5 kg), the percentage of premature births (<37 weeks of gestation), the percentage of births with late or no prenatal care (defined as starting after the first trimester of pregnancy), the infant mortality rate (number of infant deaths per 1000 live births), and the neonatal mortality rate (number of deaths to infants <28 days old per 1000 live births).[14]

Independent Variables

The two major variables of interest are policies to fund abortions for Medicaid-eligible women and policies that fund family planning services for low-income women. The measure of abortion funding policy is the funded abortion rate—that is, the number of publicly funded abortions in the state per 1000 women aged 15 to 44.[15] The per capita family planning expenditures measure includes all federal (Titles V, X, XIX, and XX) and state funds.*[16,19]

Public policies toward funding abortions and family planning are not the only factors that influence maternal and child health; numerous socioeconomic factors do also. Precise estimates of the respective impacts of abortion funding and family planning policies on maternal and child health require that these other factors be included as control variables. For example, teen pregnancy rates and proportions of children who are born into poverty (and thus have more severe health problems) are higher among non-White women. Black women are also less likely to receive adequate prenatal care.[20] In addition, Henshaw et al. found that the abortion rate is 2.7 times higher for non-White women than for White women.[21] Black women are much more likely than White women to have an unwanted birth.[7] Latino women also have higher abortion rates than non-Latino women.[22] These relationships necessitate including the Black and Latino population percentages in the model. (Data are from the Bureau of the Census, *Statistical Abstract of the United States,* for various years. For the Latino population, individual years had to be extrapolated from the 1980 and 1990 census data.)

States with fewer poor women will have fewer birth-related problems.[23] Abortion rates are also related to income; Medoff found that demand for abortions is positively related to the labor force participation of women.[24] For this reason, the percentage of women in the labor force and per capita income are also included in the model.

A proxy measure of pro-life attitudes in a state—the percentage of the state that is Catholic—is included.[25] The Catholic church has actively supported right-to-life groups and opposes abortion and most birth control methods.[15] A negative relationship between the proportion of Catholics in a state and the rate of abortions was expected.

One final control variable is access to abortion providers. Shelton et al. found that the farther a woman must travel to obtain an abortion, the less likely she is to get one.[26] The availability of abortion services varies widely across the United States; more than 30% of women of reproductive age live in counties with no abortion provider.[11] Owing to the expected positive relationship between access to providers and the abortion rate, the percentage of a state's population living in counties with large providers of abortion is included as a control variable.

Methods

The analysis is a pooled time series of all US states from 1982 to 1988. Each dependent variable was examined as the result of the six control variables as well as the funded abortion rate and family planning expenditures. A pooled time-series design is often plagued by problems of autocorrelation and heteroscedasticity.[27,28] Although the current data set is cross-sectionally dominant, preliminary analysis revealed that first-order autocorrelation was the major problem. This problem dictated that all models would be estimated using generalized least squares-autoregressive moving average models without forced homoscedasticity. After initial model estimation, residuals were examined to determine if significant fixed effects were omitted from the model. As would be expected, in many cases individual states deviated greatly from the model. To control for such influences, a series of state dummy variables were added to the equations.** The state dummys for each regression model are reported in the Appendix.

Results

The regression results for all variables appear in Table 1. Because our concern is with the impact of abortion funding and family planning expenditures, we have not interpreted the coefficients for the control variables. These coefficients are usually statistically significant and in the direction predicted by previous research.

Abortion Rates

The model predicts state abortion rates fairly well ($R^2 = .89$). Both policy impacts are as predicted: A $1 per capita increase in family planning expenditures is associated with 1.046 fewer abortions per 1000 women, consistent with the notion that effective family planning policies reduce the need for subsequent abortions. Similarly, for each additional abortion funded by a state, the number of abortions increases by 0.42.

Maternal and Child Health Indicators

The remainder of Table 1 reveals a generally good level of prediction for the child and maternal health indicators ($R^2 > .69$). There is no relationship between family planning funding and teen

*The Alan Guttmacher Institute data on funded abortions and family planning are available for 1982, 1984, 1985, and 1987. Data were extrapolated for the missing years. That is, 1986 data were the average of the 1985 and 1987 data. The 1988 data were the weighted average of 1987 and 1990 data, weighting 1987 data twice as much as 1990 data.

**The dummy variables, as expected, improve the fit of the regression line and reduce the autocorrelation. Their inclusion has little impact on the substantive interpretation of either family planning or abortion funding. The apparent impact is to reduce the size of these coefficients; the result is a conservative estimate of impact compared with the ordinary least squares estimates. For low birthweights, specific time-point dummys were also included for 1987 and 1988 to correct for time-dependence problems.

TABLE 1—The Impact of Policy on Abortions, Teen Birthrates, Low Birthweights, and Premature Births: Pooled Estimates

Independent Variables	Dependent Variables						
	Abortion Rate	Teenage Birthrate	Low Birthweight	Premature Births	Late Prenatal Care	Infant Mortality	Neonatal Mortality
Family planning funding per capita	−1.046	−.026	−0.49	.045	−.261	−.159	−.134
	(5.47)	(.11)	(2.02)	(1.03)	(1.88)	(2.29)	(2.29)
Rate of publically funded abortions	.420	−.673	−.024	−.027	−.263	−.011	.003
	(4.76)	(7.81)	(2.73)	(2.83)	(7.56)	(.70)	(.25)
Percentage Black population	.110	.385	.105	.175	.069	.138	.130
	(3.44)	(10.87)	(35.00)	(39.08)	(3.85)	(21.33)	(24.34)
Percentage Latino population	.394	.842	.022	.039	.457	−.006	−.019
	(10.06)	(25.15)	(6.01)	(9.59)	(13.09)	(.79)	(4.10)
Percentage Catholic population	−.027	−.617	−.005	−.011	−.185	.001	.018
	(1.24)	(24.86)	(2.48)	(2.96)	(13.13)	(.16)	(4.88)
Per capita income, thousands	−.217	.407	.009	.058	.052	−.191	−.158
	(3.32)	(5.41)	(.99)	(4.58)	(1.16)	(9.48)	(9.43)
Percentage female labor force participation	.076	−.255	−.059	.016	−.050	−.042	−.046
	(1.60)	(4.95)	(11.47)	(1.60)	(1.39)	(4.06)	(5.84)
Percentage population: counties with abortion facilities	.225	−.020	.003	−.006	−.003	.005	.006
	(17.57)	(2.14)	(2.79)	(3.67)	(.49)	(2.14)	(3.23)
Buse R^2	.89	.79	.79	.84	.74	.69	.69
Adjusted R^2	.88	.78	.78	.83	.72	.67	.67
Rho	.40	.50	.60	.56	.71	.26	.29

Notes. Coefficients are unstandardized regression coefficients; *t* scores are listed in parentheses. See the Appendix for fixed effects controls.

TABLE 2—National[a] Totals for Abortion Funding 1982 through 1988: Their Impact on Total Abortions and Birth Status

Indicator	States That Fund	States That Do Not Fund
Number of abortions	+563 866	−503 778
Births to teenage mothers	−151 907	+136 499
Low-birthweight births	−21 255	+18 453
Premature births	−23 928	+20 760
Late or no prenatal care	−232 570	+201 907

[a]Does not include the District of Columbia.

birthrates or premature births, but family planning funding is related to all other indicators. All other things being equal, each additional family planning dollar spent per capita by a state is associated with 0.049 percentage points fewer low-birthweight babies, 0.261 percentage points fewer births with late prenatal care, and 0.159 fewer infant deaths and 0.134 fewer neonatal deaths per 1000 live births. Funding abortions, in turn, has somewhat different impacts. All other things being equal, an increase of one funded abortion per 1000 women of childbearing age is associated with 0.673 fewer teen births per 1000 teenage women, 0.024 percentage points fewer low-birthweight babies,* 0.027 percentage points fewer premature births, and 0.263 percentage points fewer births with late or no prenatal care.

Two comments about the size of these coefficients are in order. First, funded abortions are associated with a major drop in births to teen mothers—perhaps as many as 0.67 teen births for every abortion funded. Second, many other impacts of abortion funding or family planning appear modest; however, even small changes in these variables can have a major impact.

The equations generated in Table 1 were used to estimate the impact of abortion funding and family planning expenditures in each state. To estimate the impact of abortion funding policies in states that fund abortions, the regression coefficients in Table 1 were multiplied by the actual funded abortion rate. The results were then translated from rates per 1000 women, or percentages, to actual numbers. For states that did not fund abortions, an estimate was made of what would have happened if the state *had* funded abortions. With all other state-level factors remaining the same, the mean funded abortion rate for states that actually funded abortions (4.97 per 1000 women) was used for these states. These numbers suggest the impact of not funding abortions. (The ecological analysis presented here cannot rule out the possibility that other factors caused these changes, but any such changes would have to be collinear with either funded abortions or family planning expenditures.)

Table 2 shows the national estimates. In states that funded abortions for one or more of the 7 years under analysis, there were an additional 563 900 abortions, 151 900 fewer births to teen mothers, 21 300 fewer low-birthweight babies, 23 900 fewer premature births, and 232 600 fewer births with late or no prenatal care. States that did not fund abortions had 503 800 fewer abortions, 136 500 more births to teen mothers, 18 500 more low-birthweight babies, 20 800 more premature births, and 201 900 more births

*This finding contradicts Currie et al.,[8] who find no impact. Their model used individual-level data and estimated models using least squares dummy variables. The inclusion of regional and year dummy variables might have eliminated the impact of state laws since states that fund abortions do cluster in some regions. A more precise measure of abortion funding is used here.

with late or no prenatal care. The cumulative impact of funding or not funding abortions for Medicaid-eligible women over this time period, therefore, could be substantial.

Table 3 estimates the impact of 1982–1988 family planning funding on abortions, low birthweights, late prenatal care, infant deaths, and neonatal deaths. These estimates were generated using the actual amount spent per capita in each state, and the predicted rates were then transformed into raw numbers. The total for abortions prevented, 670 000, suggests that an absence of family planning funding would likely have increased the number of abortions by about 6% nationwide. Family planning funds are also associated with 20 000 fewer low-birthweight babies and 106 900 fewer births with late or no prenatal care. These two impacts probably explain family planning's impact on infant and neonatal mortality; family planning funding was associated with 6500 fewer infant deaths and 5500 fewer neonatal deaths. Although these numbers are dwarfed by the national totals for neonatal and infant deaths, they represent a significant spillover benefit, given that family planning is directly targeted at unwanted pregnancies rather than at infant mortality.

Conclusion

Our analysis links public funding for abortion to public health benefits. States that fund abortions have substantially fewer teen births and about one-fourth fewer cases of inadequate prenatal care. Because maternal age of less than 20 years and inadequate prenatal care[29] are both risk factors for premature and low-birthweight infants,[30] it is not surprising that abortion funding is associated with a modest reduction in each of these birth outcomes.

Publicly subsidized family planning services also produce public health benefits, although the dynamics are different. Family planning funding is associated with fewer low-birthweight babies and fewer births with late or no prenatal care, and these impacts reduce neonatal and infant mortality. In contrast to abortion funding, however, family planning expenditures show almost no effect on teenage birthrates. Therefore, even though both abortion and family planning address unwanted fertility, their respective public health outcomes are distinctly different. These outcomes can be attributed to differences in clientele, delivery systems, and funding mechanisms.

Clientele

Patient characteristics may explain differences in maternal and child health outcomes. Women who have abortions are predominantly young, single, and of modest means[31]; more than half are nulliparous and have not experienced prior induced abortions.[21] Many family planning patients are also young, single, and of low income.[32] With the cutbacks in public funds for family planning during the 1980s, however, family planning clinics focused on serving existing patients. For the most part, there was no outreach to new patients and teenagers.[33] Therefore, we believe that fewer teens were served during the past decade than in previous periods. We would also surmise that a greater proportion of family planning patients have experienced a live birth or pregnancy.

Delivery Systems

Family planning services are more geographically accessible than abortion services. Publicly subsidized family planning services are delivered through provider agencies operating at 5460 sites,[34] while 90% of abortions are provided in only 1542 nonhospital facilities.[11,35] Family planning services are offered in nearly every county in the United States; on the other hand, 51% of metropolitan and 93% of nonmetropolitan counties lack an abortion provider.[11]

Family planning clinics also provide more continuous care than do most abortion providers. When a woman receives services at a family planning clinic, she is expected to be a continuing patient. Moreover, a family planning patient has entered a health care delivery system where she will be provided with or referred to other maternal and child health services as needed. In contrast, a woman who receives an abortion is far more likely to get only a single health service. This is because most abortions are performed in clinics where abortion is the main service provided,[33] and referral patterns are unlikely to be as well established as they are in a family planning clinic. An abortion patient is also more likely to live in another county or state, which also decreases the likelihood of a successful referral to another health care provider.

TABLE 3—National Totals for Benefits from Family Planning

Indicator	50-State Total
Decline in low-birthweight births	20 025
Decline in late prenatal care	106 867
Decline in infant deaths	6 498
Decline in neonate deaths	5 476
Decline in number of abortions	670 384

Funding Mechanisms

A third difference between abortion and family planning services is how they are financed. The 12 states that fund abortions do so through state Medicaid programs, a fee-for-service funding mechanism whereby patients go to a health care provider for a single service. In contrast, even though Medicaid is now the largest single contributor to the national family planning effort, most family planning providers receive funds from a variety of public sources.[19] If the family planning agency receives *any* Title X funds, it must provide an array of services including appropriate referrals for other health services.

The difference in the public health outcomes of abortion and family planning services presents a strong argument for a more integrated approach to the delivery of both services. The fact that so many young women initiate family planning *after* an unwanted pregnancy certainly calls for a greater and more effective family planning effort. Given the current realities of service delivery and funding, the projected outcomes in Tables 2 and 3 provide reasonable data for policy deliberations concerning family planning and abortion funding.

States seldom make policy on abortion funding based on public health benefits. The recent congressional debate on abortion funding indicates that this holds true for the federal government as well. However, we hope that this information will at least contribute to the debate when it inevitably recurs. □

References

1. Legge JS. *Abortion Policy*. Albany, NY: SUNY Press; 1985.
2. Alan Guttmacher Institute. House vote on Hyde Amendment results in only slight

modification. *Washington Memo.* July 1, 1993; Wm-11:1–2.
3. Grossman M, Jacobowitz S. Variations in infant mortality rates among counties of the United States. *Demography.* 1981;18: 695–713.
4. Corman H, Grossman M. Determinants of neonatal mortality rates in the US. *J Health Econ.* 1985;4:213–236.
5. Joyce TJ, Grossman M. Pregnancy wantedness and the early initiation of prenatal care. *Demography.* 1990;27:1–17.
6. Weller RH, Eberstein IW, Bailey M. Pregnancy wantedness and maternal behavior during pregnancy. *Demography.* 1987;24: 407–412.
7. Marsiglio W, Mott FL. Does wanting to become pregnant with a first child affect subsequent maternal behaviors and infant birth weight? *J Marriage Fam.* 1988;50:1023–1036.
8. Currie J, Nixon L, Cole N. Restrictions of Medicaid funding of abortion. Cambridge, Mass: National Bureau of Economic Research; 1993. NBER Working paper 4432.
9. Torres A, Donovan P, Dittes N, Forrest JD. Public benefits and costs of government funding for abortion. *Fam Plann Perspect.* 1986;18:111–118.
10. Alan Guttmacher Institute. Hill attention on FOCA, FACE increases following murder of Florida physician. *Washington Memo.* April 14, 1993;Wm-6:2–3.
11. Henshaw SK, Van Vort J. Abortion services in the United States, 1987–1988. *Fam Plann Perspect.* 1990;22:102–108, 142.
12. Henshaw SK, Van Vort J. *Abortion Services in the United States, Each State and Metropolitan Area,* 1984–1985. New York, NY: Alan Guttmacher Institute; 1988.
13. McFarlane DR, Meier KJ. Determinants of abortion levels in the American states, 1982–1988. Presented at the 120th Annual Meeting of the American Public Health Association; November 8–12, 1992; Washington, DC.
14. *Vital Statistics of the United States.* Hyattsville, Md: National Center for Health Statistics; 1982–1988.
15. Meier KJ, McFarlane DR. The politics of funding abortion: state responses to the political environment. *Am Polit Q.* 1993;21: 81–101.
16. Gold RB, Macias J. Public funding of contraceptive, sterilization, and abortion services, 1985. *Fam Plann Perspect.* 1986;18: 259–264.
17. Gold RB, Guardado S. Public funding of family planning, sterilization and abortion services. *Fam Plann Perspect.* 1988;20:228–233.
18. Gold RB, Nestor B. Public funding of contraceptive, sterilization, and abortion services, 1983. *Fam Plann Perspect.* 1985;17: 25–30.
19. Gold RB, Daley D. Public funding of contraceptive, sterilization, and abortion services, fiscal year 1990. *Fam Plann Perspect.* 1991;23:204–211.
20. National Center for Health Statistics. Advance report of final natality statistics, 1991. *Month Vital Stat Rep.* 1993;42(3): Tables 30 and 31.
21. Henshaw SK, Koonin LM, Smith JC. Characteristics of US women having abortions, 1987. *Fam Plann Perspect.* 1991;23: 75.
22. Henshaw SK, Silverman J. The characteristics and prior contraceptive use of US abortion patients. *Fam Plann Perspect.* 1990;22:162.
23. Joyce TJ. The impact of induced abortion on black and white birth outcomes in the United States. *Demography.* 1987;24:229–244.
24. Medoff MH. An economic analysis of the demand for abortions. *Econ Inquiry.* 1988; 26:353–359.
25. Quinn BH, Anderson H, Bradley M, Goetting P, Shriver P. *Churches and Church Membership 1980.* Atlanta, Ga: Glenmary Research Center; 1982.
26. Shelton J, Brann EA, Shultz KF. Abortion utilization: does travel distance matter? *Fam Plann Perspect.* 1976;8:260–262.
27. Hsiao C. *Analysis of Panel Data.* New York, NY: Cambridge University Press; 1986.
28. Pindyck RS, Rubinfeld DL. *Econometric Models and Econometric Forecasts.* New York, NY: McGraw-Hill; 1991.
29. Singh SS, Torres A, Forrest JD. The need for prenatal care in the United States. *Fam Plann Perspect.* 1985;17:118–124.
30. Brown SS. Can low birth weight be prevented? *Fam Plann Perspect.* 1985;17: 112–118.
31. Gold RB. *Abortion and Women's Health.* New York, NY: Alan Guttmacher Institute; 1990.
32. Alan Guttmacher Institute. *Current Functioning and Future Priorities in Family Planning Services Delivery.* New York, NY: Alan Guttmacher Institute; 1990.
33. Dryfoos JG. What President Bush can do about family planning. *Am J Public Health.* 1989;79:689–690.
34. Henshaw SK, Torres A. Family planning agencies: services, policies, and funding. *Fam Plann Perspect.* 1994;26:52–59.
35. Henshaw SK. The accessibility of abortion services in the United States. *Fam Plann Perspect.* 1991;23:247.

APPENDIX—Fixed Effects Estimated for the Following States in the Equations Listed in Table 1

Abortion rate: Arkansas, Connecticut, Delaware, Florida, Georgia, Idaho, Indiana, Kentucky, Maine, Maryland, Massachusetts, Mississippi, Nevada, New Mexico, North Carolina, Rhode Island, South Carolina, South Dakota, Utah, West Virginia, Wyoming.

Teen birthrates: Colorado, Iowa, Kentucky, Louisiana, Maryland, Minnesota, Mississippi, New York, Oklahoma, Rhode Island, Utah, Virginia, Washington.

Low birthweights: Alaska, Colorado, Hawaii, Nevada, Washington, Wyoming.

Premature births: Connecticut, Hawaii, Kentucky, Maryland, Minnesota, Rhode Island, Washington, West Virginia, Wyoming.

Late prenatal care: Arkansas, Iowa, Minnesota, New Mexico, New York, Oklahoma, South Carolina, South Dakota, Utah, Virginia, West Virginia.

Infant mortality: Alaska, Arkansas, Illinois, Louisiana, South Dakota, Wyoming.

Neonatal mortality: Alabama, Arkansas, Delaware, Illinois, Louisiana, Mississippi, New Hampshire, Washington.

The Costs and Financing of Perinatal Care in the United States

ABSTRACT

Objectives. The purpose of this study was to estimate the aggregate annual costs of maternal and infant health care and to describe the flow of funds that finance that care.

Methods. Estimates of costs and financing based on household and provider surveys, third-party claims data, and hospital discharge data were combined into a single, best estimate.

Results. The total cost of perinatal care in 1989 was $27.8 billion, or $6850 per mother–infant pair. Payments made directly by patients or third parties for this care totaled $25.4 billion, or about 7% of personal health care spending by the nonaged population. Payments were less than costs because they did not include a value for direct delivery care or for bad debt and charity care, which accounted for $2.4 billion. Private insurance accounted for about 63% of total payments, and Medicaid accounted for 17% of the total.

Conclusions. National health reform would provide windfall receipts to hospitals, which would receive payment for the considerable bad debt and charity care they provide. Reform might also provide short-term gains to providers as private payment rates are substituted for those of Medicaid. (*Am J Public Health.* 1994;84:1473–1478)

Stephen H. Long, PhD, M. Susan Marquis, PhD, and Ellen R. Harrison, MA

Introduction

The financing, delivery, and medical effectiveness of perinatal care in the United States has been a subject of continuing concern for decades. Most obvious among the causes of concern are the poor relative position of the US infant mortality rate among those of other nations and the persistent disparity in infant mortality rates between Blacks and Whites in this country.[1,2] A number of distinguished commissions and panels have studied the problem,[3–5] and major restructuring of the delivery system has been implemented.[6] Recently, significant expansions of eligibility for Medicaid financed care have been made for pregnant women and infants.[7] The prospect of national health reform heralds future changes in spending and financing for maternal and infant health care services.

One dimension that needs to be monitored as the nation continues to wrestle with this problem is the financing of perinatal care. That is, how much do we spend? For what services? Who pays the bills? How will reforms alter the mix? Several previous studies have attempted to answer one or more of these questions.

Since the 1970s, the Health Insurance Association of America has produced periodic estimates of typical charges for inpatient hospital care and fees for physician services associated with a normal delivery and a cesarean section.[8,9] Weighting the association's most recent (1989) estimates of the typical charge for each procedure by their respective frequencies, the per capita charges for the specified maternity care services were about $5050.

Fuchs and Perreault provided the first estimates of aggregate expenditures on reproduction-related health care—including obstetrical care, infant care, contraception, abortion, and infertility services—in the United States.[10] Aggregate obstetrical spending was based on the Health Insurance Association of America estimates of per capita charges for hospital and physician services multiplied by the total number of births, supplemented by estimates of additional tests not accounted for in the association's service bundle. Spending on infant care was estimated from Health Insurance Association of America charge data on nursery care for newborns and estimates of the use and charges for neonatal intensive care, well baby care, and other services. Inflating the Fuchs and Perreault estimates for 1982 to 1989 dollars—using the medical care component of the Consumer Price Index—and births, aggregate spending for obstetrical and infant care would be $14.5 billion and $11.5 billion, respectively.

A series of reports by staff of the Alan Guttmacher Institute assembled estimates of charges for perinatal care and provided considerable information about the major sources of third-party financing for such care.[11–14] For inpatient hospital charges, the authors used either the above Health Insurance Association of America survey or a survey conducted by the Commission for Professional Hospital Activities. Physician charges were also taken from the Health Insurance Association of America. Multiplying these estimated charges by the number of births, the institute's 1985 estimate of total charges for maternity care was $23 billion, given 1989 dollars and births.

The authors are with RAND, Washington, DC.

Requests for reprints should be sent to Stephen H. Long, PhD, RAND, 2100 M St, NW, Washington, DC 20037.

This paper was accepted May 25, 1994.

Editor's Note. See related editorial by McCormick and Lee (p 1376) in this issue.

Recently, Lewit and Monheit updated the Fuchs and Perreault estimates of the average cost and aggregate cost of US maternity and infant care.[15] Again, they relied on the Health Insurance Association of America estimates of hospital and physician charges to value prenatal, delivery, and routine newborn care. For infant care, they used recent data from the 1987 National Medical Expenditure Survey (NMES). Aggregate expenditures on perinatal care in 1987, adjusted to 1989 dollars and births, were $34 billion according to this source.

The purpose of the present study was to build and improve on this earlier literature in several ways. First, previous work has used charges as the measure of valuing services. Since few third-party payers actually pay full charges, and because maternity care is a well-documented source of bad debt and charity care, we introduce the concepts of payment and cost as alternative measures of value. Second, earlier studies often have depended on surveys to measure charges for the "typical" or "modal" service bundle. Aggregate totals produced from such estimates are at risk of considerable error, should either the quantity of use or the charge for the modal service mix differ from their averages. To measure aggregate spending more accurately, we used administrative records for episodes of care whenever possible to measure the "mean" charge, thereby accounting for outliers, whether they be low users of prenatal care or extremely high-cost hospitalizations. Third, we attempted to disaggregate services into discrete episodes of care for mothers and infants somewhat more extensively than has been done in the past. Fourth, we examined the distribution of spending by source of payment as well as by type of spending; except for the Alan Guttmacher Institute study, previous studies have looked only at the distribution by service type. Finally, we greatly expanded the number of sources used to make the estimates, contributing, we hope, to their accuracy.

Here we measure costs and financing of maternal and infant care in 1989. We chose that year because it preceded most of the recent Medicaid eligibility expansions and provided a baseline against which those expansions and other reforms could be monitored. We conclude with a discussion of the implications of Medicaid reforms and national health reforms for the mix of spending and financing observed in 1989.

Methods

The Flow of Funds Matrix Framework

The structure that we used to measure the financing of maternal and infant health care in the United States was modeled after the National Health Expenditures Accounts, a continuing series currently maintained by the Health Care Financing Administration to describe the flow of funds financing all health care in the United States.[16] The flow of funds matrix is a way of describing activities; it allowed us to know the magnitude of the current system expressed in the common metric of dollars.

The columns of our flow of funds matrix categorize expenditures by type of service, and the rows categorize expenditures by payer. The services that we measured included maternity care from conception until 60 days after delivery and infant care through the first year of life. Although the term "perinatal care," in its technical usage, would not include all of the infant care we measured, we use "perinatal" as a shorthand in this article to refer to all of these maternal and infant services. The service categories classify this care by the stage of the process: prenatal and postdelivery care, delivery services (those provided to the mother at delivery), newborn care (services provided to the newborn from birth to the initial hospital discharge), and infant care (care delivered from hospital discharge through the first year of life). These categories are further classified by provider (inpatient hospital services and services of physicians and other health care providers).

The rows of the matrix categorize expenditures by payer. Some of the care is covered by private third-party payers and some by Medicaid and other government insurance programs. The latter category includes CHAMPUS (the Civilian Health and Medical Program of the Uniformed Services), other care paid for by the Department of Defense, and other state and federal programs that make payments on behalf of a patient receiving care. Self-pay includes patient payments made by uninsured patients; for insured patients, it includes their payments for cost sharing and services not covered by their plans. The final category is direct delivery care, within which we include all care provided to a patient without payment on behalf of that particular patient. This includes care financed by government grants and contracts, such as first provided by community and migrant health centers, Title V clinics, and Crippled Children's Services, as well as bad debt and charity care from physicians and hospitals.

Measurement Concepts

It was noted above that, when estimating expenditures, all earlier studies have valued the resources devoted to perinatal care according to the charges of hospitals and the prevailing fees charged by physicians. Increasingly over the last decade, third-party payers have either refused to pay full charges, as has Medicaid, or negotiated discounts, as have Blue Cross and Blue Shield and most managed care organizations. Moreover, uninsured patients are frequently unable to pay much, if any, of a hospital's normal charge for delivery, but generally they are not denied care. Therefore, using charges to estimate direct expenditures almost surely results in an overestimate.

Instead of charges, we used two alternative measures to value perinatal care resources. First, we estimated direct payments for care by patients and on account of patients by third-party payers. In other words, we attempted to measure what was actually collected by the provider for the care of a particular patient. This concept does not include contributions that are not tied to particular patients, such as federal block grants to states for Title V programs and general contributions by local governments to public hospitals for charity care. Second, we estimated the cost of care provided to patients. In the case of hospital services, this is the value of the resources the hospital assembles to provide care, including wages and salaries, supplies, utilities, and depreciation on plant and equipment. In the case of physician services, this represents practice costs plus the monetary return physicians accept to see their patients, whether paying full fees or not.

The payment may be either greater or less than the cost. For example, private insurers pay most or all of the provider's charge, which generally exceeds the average cost of providing the service. On the other hand, our payment concept attaches no dollar value to government direct delivery services or charity care to a patient because no direct payment for the patient's care is involved. In contrast, our cost concept places the same positive cost value on a unit of care delivered in a public health clinic or on a unit of charity care as it does on a unit of care for which the full charge was paid.

Data and Estimation Methods

Overview. Our estimates of aggregate payments or costs for each cell of the matrix were based on a variety of different data sources, including household and provider surveys, third-party claims data, hospital discharge data, and the published literature. From these sources, we estimated the average charge per delivery or per infant for each of our service categories (the column totals) and multiplied these per case estimates by data from the vital statistics system on the number of births and fetal and infant deaths to arrive at the aggregate service totals. We allocated the aggregate service totals to payer using estimates of the distribution of charges for each of the service categories (or columns) by payer from our data sources. We converted the resulting matrix of charges to payments and costs using estimates of the ratio of payments to charges for different payers and of costs to charges.

Data availability dictated the estimation strategy that we adopted. There was no data source that allowed us to directly estimate any cell in our payment or cost matrix and only one data source that permitted us to estimate charges categorized by both type of services and type of payer for all maternal and infant spending. Some data sources provide information only about spending in one setting (e.g., inpatient care) or for selected stages of care (e.g., delivery services). Other sources provide estimates of total spending but do not categorize it by type of service or payer. The single source that provides information about charges by type of service and payer is a sample survey, and some cell estimates derived from this database are subject to respondent reporting error and are unreliable because of very small sample size. As a result, we have pooled a number of data sources to estimate our matrix.

Data on charges per case. The data sources included in our estimate of charges per case were NMES, a 1987 survey of about 14 000 US households; 1990 insurance claims data for patients covered by the Aetna Insurance Company; 1989 data from the MEDSTAT data system representing the health care use experience of more than half a million workers enrolled in one of 32 different group health insurance plans in all areas of the country; the Health Insurance Experiment, a controlled trial in health care financing that collected health insurance claims data from almost 6000 individuals in six sites throughout the country for a 3- or 5-year period[17]; 3 years of insurance claims data for employees and dependents of three large employers in two states; Medicaid claims data from four states[18–20]; the California Office of Statewide Health Planning and Development 1988 Annual Discharge Abstracts, which contain a record for each discharge from a California hospital in the year; data from the Hospital Cost and Utilization Project, a source of hospital discharge abstracts from a national sample of hospitals; and published estimates from the Health Insurance Association of America[9] and the Alan Guttmacher Institute.[13]

Our data sources came from different years and, in some cases, from limited geographic areas. We adjusted all dollar estimates to 1989 dollars using the hospital or physician services component of the Consumer Price Index. We adjusted per case expenditure estimates based on data from a single state to national prices using the Medicare Geographic Practice Cost Index.

We averaged the estimates of per case charges from the various databases to obtain a best estimate. Because we gave equal weight to each database in computing our estimate without regard to the payer group represented (e.g., patients covered by private payers vs Medicaid patients), we adopted the assumption that average patient needs and the average service mix do not vary by payer status. Although this may impart some bias, the direction is not clear. For example, Howell et al.[19] found that women on Medicaid tend to initiate prenatal care later in the pregnancy and are less likely to have a cesarean section delivery, both of which would lead to lower charges. In contrast, women on Medicaid have poorer birth outcomes than non-Medicaid women, probably leading to higher charges for infant care.

We had more confidence in estimates from the databases of the average total charge for maternity care and newborn care at delivery than for the allocation between prenatal care, care for the mother at delivery, and newborn care because of differences across payers and providers in billing practices. For example, some physicians submit global bills for all maternity care, whereas some submit separate bills for each service. Some hospitals have a single bill for both the newborn and the mother, whereas others submit separate bills. Moreover, there is less variability among studies in the estimate of average total charges than in estimates by type of service. We also had more information about total charges for the mother and newborn or charges during the first year of life than we had about the components, because some databases do not provide the information to allocate charges by type of provider or type of service. Therefore, we first estimated per case total charges for the mother and newborn at delivery and per case total charges for the first year of life by averaging the estimates from our various data sources. Then we distributed these charges among service and provider categories using the average share of the total attributable to the category estimated from the different databases. The service-specific per case charge estimates were multiplied by data on the number of cases from the vital statistics system to produce aggregate charge estimates. An appendix, available on request from the authors, provides more detail about the estimation and summarizes the data used to construct the charge matrix. This appendix also reports some sensitivity analyses that show our estimates vary by only about 10% when we alter some key assumptions, and it presents some independent aggregate estimates where available for specific payer and service categories that validate our constructed estimates.

Distributing charges among payers. Aggregate charges were distributed among the payers by using estimates of the service-specific share of the total attributable to the payer. Estimates of the payer share came from the following databases: the California Hospital Discharge database; the Hospital Cost and Utilization Project database; NMES; the National Hospital Discharge Survey, a sample of discharge records from about 400 hospitals nationwide; the 1988 National Maternal and Infant Health Survey, a retrospective history of the use patterns of more than 18 000 pregnant women; and the National Ambulatory Medical Care Survey, a sample of patient records from a nationwide sample of physicians. The data used in constructing payer shares are also available from the authors by requesting the appendix.

Calculating payments and costs from charges. Aggregate inpatient hospital charges for each payer were adjusted to reflect payments by means of payment-to-charge ratios by payer category derived from data from the American Hospital Association,[21] the Hospital Cost and Utilization Project (for private and public third parties), and a large private insurer (for private third parties). The payment-to-charge ratio is always less than one,

TABLE 1—Flow of Patient Payments for US Maternal and Infant Health Care, by Source and Use of Funds, 1989

Payer	Maternity Prenatal and Postdelivery Services[a]		Delivery Services		Newborn Care[b]		Infant Care[c]		Total
	Inpatient Hospital	Physician and Other Services	Inpatient Hospital	Physician and Other Services	Inpatient Hospital	Physician and Other Services	Inpatient Hospital	Physician and Other Services	
Payments, billions of dollars									
Private insurance	0.48	2.01	5.27	3.11	1.80	0.79	1.31	1.33	16.10
Medicaid	0.18	0.32	1.66	0.42	0.64	0.14	0.59	0.26	4.21
Other government insurance	0.05	0.15	0.59	0.08	0.18	0.12	0.20	0.36	1.72
Self-pay	0.08	0.48	0.72	0.46	0.29	0.24	0.25	0.82	3.35
Total	0.79	2.95	8.24	4.07	2.90	1.29	2.35	2.78	25.37
Per case payment, in dollars	190	720	2020	1000	720	320	590	690	6250
Distribution by payer, %[d]									
Private insurance	3	13	30	19	11	5	8	8	100
Medicaid	4	8	40	10	15	3	14	6	100
Other government insurance	3	9	34	5	10	7	11	21	100
Self-pay	2	14	22	14	9	7	8	24	100
Total	3	12	33	16	11	5	9	11	100
Distribution by service, %[e]									
Private insurance	61	68	64	76	62	61	56	48	63
Medicaid	23	11	20	11	22	11	25	10	16
Other government insurance	6	5	7	2	6	9	8	13	7
Self-pay	10	16	9	11	10	19	11	29	13
Total	100	100	100	100	100	100	100	100	100

Source. The totals for and most of the data under "Payments, billions of dollars" are reprinted with permission from The March of Dimes.[24]
[a]Includes care delivered up to 60 days postdelivery.
[b]Includes care delivered from birth to initial hospital discharge.
[c]Includes care delivered from hospital discharge through first year of life.
[d]Row percentages.
[e]Column percentages.

reflecting deductions from gross charges for contractual adjustments (or "discounts") and uncompensated care for the pool of patients having the same primary payer type. For physician services, we based our estimates of payment-to-charge ratios on data from the American Medical Association's (AMA) Socioeconomic Monitoring System, which is an ongoing survey of nearly 3000 physicians.[22,23] Specifically, we calculated these ratios on the basis of data on discounts from the usual fee for established patient office visits by payer and specialty-specific data on charity care provided, as a percentage of practice time, for obstetrics/gynecology, general/family practice, and pediatrics. For private third-party payers, we averaged the AMA estimate of the payment-to-charge ratio with one obtained from a large private insurer.

Hospital costs were estimated on the basis of an overall hospital cost-to-charge ratio from the Hospital Cost and Utilization Project; that is, we implicitly assumed that costs per patient did not vary by payer type. Our estimate of the cost-to-charge ratio for physician services was the ratio of our estimate of aggregate payments to aggregate charges net of charges for direct delivery care. In effect, it was a cost-to-charge ratio based on the observation that, after covering office expenses, this was the compensation for which private practice physicians were willing to work to provide this total bundle of services, whether paid or not. Costs for each cell were calculated by multiplying charges by the cost-to-charge ratio. This same ratio was applied to our estimate of charges for direct delivery care.

Results

Payments for Maternity and Infant Care in 1989

Payments for maternity and infant care were estimated to be $25.4 billion, or about $6250 per mother–infant pair, in 1989 (Table 1). These payments included those made directly by patients to providers or made by a third party on behalf of a patient; they did not include a value for direct delivery care or bad debt and charity care. Such payments represented about 4.5% of personal health care spending by all Americans and about 7% of health care spending by the nonelderly population.

About 15% of total payments were for prenatal and postdelivery services ($910 per mother), 48% were for obstetric delivery services ($3020 per mother), 17% were for newborn care ($1040 per newborn), and 20% were payments for care received during the first year of life ($1280 per infant).

Payments to hospitals accounted for 56% of the total, and payments to physicians and other providers accounted for 44%. Hospital payments on behalf of the mother totaled $9 billion and physician payments $7 billion. Hospital payments for newborns and infants during the first year of life were $5.3 billion, and

TABLE 2—Total Cost of US Maternal and Infant Health Care, by Payer and Type of Service, 1989

Payer	Maternity Prenatal and Postdelivery Services[a]		Delivery Services		Newborn Care[b]		Infant Care[c]		Total
	Inpatient Hospital	Physician and Other Services	Inpatient Hospital	Physician and Other Services	Inpatient Hospital	Physician and Other Services	Inpatient Hospital	Physician and Other Services	
Costs, billions of dollars									
Private insurance	0.42	1.75	4.61	2.78	1.59	0.67	1.11	1.09	14.02
Medicaid	0.23	0.44	2.16	0.61	0.84	0.19	0.74	0.35	5.56
Other government insurance	0.05	0.17	0.62	0.09	0.19	0.13	0.20	0.38	1.83
Self-pay	0.09	0.60	0.86	0.58	0.29	0.30	0.29	0.96	3.97
Direct delivery and uncompensated care	0.10	0.08	0.85	0.61	0.34	0.04	0.33	0.07	2.42
Total	0.89	3.03	9.09	4.68	3.25	1.33	2.68	2.85	27.79
Per case cost, in dollars	220	740	2230	1150	800	330	670	710	6850
Distribution by payer, %[d]									
Private insurance	3	12	33	20	11	5	8	8	100
Medicaid	4	8	39	11	15	4	13	6	100
Other government insurance	3	9	33	5	11	7	11	21	100
Self-pay	2	15	22	15	7	8	7	24	100
Direct delivery and uncompensated care	4	3	35	25	14	2	14	3	100
Total	3	11	33	17	12	5	9	10	100
Distribution by service, %[e]									
Private insurance	47	58	51	59	49	50	41	38	50
Medicaid	26	15	24	13	26	14	28	12	20
Other government insurance	6	5	7	2	6	10	8	13	7
Self-pay	10	20	9	13	9	23	11	34	14
Direct delivery and uncompensated care	11	2	9	13	10	3	12	3	9
Total	100	100	100	100	100	100	100	100	100

Source. The totals for and most of the data under "Costs, billions of dollars" are reprinted with permission from The March of Dimes.[24]

[a]Includes care delivered up to 60 days postdelivery.
[b]Includes care delivered from birth to initial hospital discharge.
[c]Includes care delivered from hospital discharge through first year of life.
[d]Row percentages.
[e]Column percentages.

payments of other providers totaled $4.1 billion.

Private insurance paid about 63% of the total; Medicaid, 17%; other government insurance programs, 7%; and out-of-pocket payments by patients, 14%. The Medicaid share, which totaled $4.2 billion, represented 7% of Medicaid outlays in 1989.

Private insurance paid a smaller fraction of infant care than of maternity and newborn care (52% vs 67%) with out-of-pocket payments picking up a larger fraction of the infant care (21% vs 11%). Medicaid accounted for a somewhat smaller fraction of physician payments than of hospital payments (10% vs 21%) because Medicaid discounts relative to private payer discounts were greater for physician than for hospital services.

Costs of Maternity and Infant Care in 1989

The total cost of maternity and infant health care in 1989 was $27.8 billion, or $6850 per mother–infant pair (Table 2). This exceeded payments made directly on behalf of patients by $2.4 billion, which represented uncompensated care and the direct delivery of maternal and infant health services.

The cost of services covered by private payers was less than the payments actually made by these third parties; payments exceeded costs by almost $2 billion, or 13%. This was due in part to cost shifting by providers to private payers to collect for charity care and for care to Medicaid patients that, on average, was reimbursed at less than cost. As a result, the cost of services covered by public insurance programs or paid for directly by the patient exceeded the payments made for these services. For example, Medicaid paid $4.2 billion for perinatal care services, whereas we estimated the cost of these services to be $5.6 billion.

The value of uncompensated hospital care and hospital care provided through direct delivery programs was $1.6 billion, or 10% of the total cost of hospital services for maternity and infant care. For physician services, the value of direct delivery services and other uncompensated services amounted to $0.8 billion, or 7% of the total cost of physician services for pregnant women and infants.

Discussion

How is the level of perinatal care financing and its pattern by payer and by

service that we measured for 1989 likely to change in the future? Many trends will affect the level and pattern of financing, including technological changes in the treatment of high-cost infants and changing practice patterns, such as shorter inpatient hospital stays for normal deliveries. Although we cannot comment on all important trends here, we offer some thoughts about the effects of two forces: the recent Medicaid eligibility expansions for pregnant women and infants and the prospect of national health reform resulting in universal health insurance coverage for all Americans.

The share of infants covered by Medicaid increased substantially between 1989 and 1992 during the phasing in of the Medicaid eligibility expansion, from 22% to 31%, or 9 percentage points (based on our tabulations of the Current Population Survey). What is the implication for provider payments? If the newly eligible were formerly uninsured, then providers should be experiencing some revenue gains, since the costs of treating these women and newborns are not likely to have changed. But the proportion of uninsured infants fell by only 2 percentage points, from 14% to 12%. Nearly all of the remainder of the shift among payers came through a 6 percentage point reduction in private insurance coverage. If these estimates prove accurate, and given the relative payment rates of Medicaid and private insurance, expanding Medicaid into higher ranges of the income distribution may not deliver the effect on the bottom line once envisioned by some provider groups.

The effect of national health reform on the quantity and pattern of current financing of perinatal care would depend on several factors, including the minimum benefit package (if any), changes in payment rates by payers, and behavioral changes by providers and by patients. The use of inpatient hospital services for delivery and newborn care seems unlikely to change much in response to health reform. Although these services would surely be included in any minimum benefit package, the uninsured tend to receive care in the current system despite their inability to pay for it. An increase in use of prenatal and well baby care under universal insurance seems more likely, should they be included in a basic benefit package. The different education and income levels of the uninsured relative to those of the insured make the magnitude of the increase quite uncertain, however.

In contrast, the effects on the pattern of financing would be substantial. Universal insurance might provide substantial windfall receipts to hospitals, at least in the short run. At minimum, they would receive payment for nearly all of the considerable bad debt and charity care they now provide. In addition, many health reform plans would replace Medicaid acute care with government-subsidized private insurance. As we observed earlier, private payments now exceed the cost of services provided to patients, in part reflecting shifting costs of public patients whose care is reimbursed below cost. Thus, as private payment rates are substituted for those of Medicaid, hospitals will almost surely realize windfall profits.

The ultimate financial outcome, however, would depend on the net effect of changes in payments from various of their other payers. For example, under managed competition plans these short-term windfalls might be given away by hospitals over the longer run in the form of discounts to cost-conscious health plans. Alternatively, under a more regulated model, the government might take them back, either by regulating lower payments from other payers or by imposing a hospital excise tax. The implication of such changes is that our two measures of value, payments and costs, are likely to come much closer to resembling one another, and the current variation in payment-to-cost ratios by payer type is likely to be reduced or eliminated. Under some reform models, this is a tautology, since they would either greatly reduce the number of payers or reduce their freedom to use their own payment amounts.

Aside from the possibility of larger increases in the demand for their services under national health reform, the implications of payment changes for physicians are similar to those for hospitals. Whatever bad debt and charity care they would otherwise provide would become paid care if everyone were insured for their services, aside from any cost sharing in the plan. And Medicaid might become better paid. But physicians' ability to sustain the windfall over the long run would depend on the decisions of other payers, government regulation and taxation, and the competitive behavior of their colleagues. □

References

1. *Health, United States, 1990.* Hyattsville, Md: National Center for Health Statistics; 1991.
2. National Center for Health Statistics. Births, marriages, divorces, and deaths for 1990. *Month Vital Stat Rep.* 1991; 39(13).
3. Institute of Medicine. *Preventing Low Birthweight.* Washington, DC: National Academy Press; 1985.
4. Institute of Medicine. *Prenatal Care: Reaching Mothers, Reaching Infants.* Washington, DC: National Academy Press; 1988.
5. *Death before Life: The Tragedy of Infant Mortality.* Washington, DC: National Commission to Prevent Infant Mortality; 1988.
6. Committee on Perinatal Health. *Toward Improving the Outcome of Pregnancy: Recommendations for the Regional Development of Maternal and Perinatal Health Services.* White Plains, NY: The National Foundation–March of Dimes; 1976.
7. Hill T. *The Medicaid Expansions for Pregnant Women and Children: A State Program Characteristics Information Base.* Washington, DC: Health Systems Research Inc; 1992.
8. *The Cost of Having a Baby.* Washington, DC: Health Insurance Association of America; 1982.
9. Minor AF. *The Cost of Maternity Care and Childbirth in the United States, 1989.* Washington, DC: Health Insurance Association of America; 1989.
10. Fuchs VR, Perreault L. Expenditures for reproduction-related health care. *JAMA.* 1986;225:76–81.
11. Gold RB, Kenney AM. Paying for maternity care. *Fam Plann Perspect.* 1985;17:103–111.
12. Gold RB, Kenney AM, Singh S. Paying for maternity care in the United States. *Fam Plann Perspect.* 1987;19:190–206.
13. *Blessed Events and the Bottom Line: Financing Maternity Care in the U.S.* New York, NY: Alan Guttmacher Institute; 1987.
14. *The Financing of Maternity Care in the United States.* New York, NY: Alan Guttmacher Institute; 1987.
15. Lewit EM, Monheit AC. Expenditures on health care for children and pregnant women. *Future Child.* 1992;2:95–114.
16. Levit K, Lazenby H, Cowan C, Letsch S. National health expenditures, 1990. *Health Care Financing Rev.* 1991;13:29–54.
17. Manning WG, et al. Health insurance and the demand for medical care. *Am Econ Rev.* 1987;77:251–277.
18. Howell E, Brown G. Prenatal, delivery, and infant care under Medicaid in three states. *Health Care Financing Rev.* 1989;10:1–16.
19. Howell E, Herz E, Wang R, Hirsch M. A comparison of Medicaid and non-Medicaid obstetrical care in California. *Health Care Financing Rev.* 1991;12:1–16.
20. Rymer M, Adler G. Children and Medicaid: the experience in four states. *Health Care Financing Rev.* 1987;9:1–20.
21. *1990 AHA Annual Survey.* Chicago, Ill: American Hospital Association; 1990.
22. Moser JW. Third-party reimbursement and physician fees. In: Gonzalez ML, Emmons DW, eds. *Socioeconomic Characteristics of Medical Practice.* Chicago, Ill: American Medical Association; 1987.
23. Physician provision of charity care. *Physician Marketplace Update* (Chicago, Ill: American Medical Association, Center for Health Policy Research). 1992(3).
24. *Toward Improving the Outcome of Pregnancy: The 90s and Beyond.* White Plains, NY: March of Dimes Birth Defects Foundation, Committee on Perinatal Health; 1993.

Public Health Briefs

Targeting Cancer Control: The State Cancer Control Map and Data Program

Forrest A. Pommerenke, MD, Robert W. Miller, MD, DrPH, Sudhir Srivastava, PhD, and Susan P. Ackermann, PhD

ABSTRACT

The State Cancer Control Map and Data Program (National Technical Information Service, Springfield, Va) allows state and county mortality maps and data tables to be tailored according to the user's choice of cancer site, age, gender, race, county, and time period between 1953 and 1987. Counties and subpopulations within these maps and data tables that are identified as having exceptionally high rates of specific cancers can then be targeted for early detection or primary prevention. These maps and data tables provide a means for focusing state, local, and individual efforts to reduce cancer mortality in appropriate populations in areas with high mortality rates. (*Am J Public Health.* 1994;84:1479–1482)

Introduction

A substantial reduction in cancer mortality can come from improvements in early detection by primary care physicians and from efforts to decrease socioeconomic, environmental, and behavioral risk factors, especially tobacco use.[1,2] Cancer control should apply to all persons but especially to populations who by age, gender, and race are at high risk for type-specific cancers.[3]

Practicing physicians and community health officials may be at a loss as to where and how to begin cancer control programs. One starting place is the examination of local cancer mortality patterns.[4] The first *Atlas of Cancer Mortality,* developed in the United States by Fraumeni and his associates, provided a panoramic view of variations in cancer mortality by county.[5] It was immensely useful for studies of etiology[6] and was followed by two atlases of cancer mortality among US non-Whites[7,8] and another of cancer mortality among Whites[9] by aggregates of counties known as State Economic Areas. The State Cancer Control Map and Data Program (hereafter called the Map Program) described here, a variation on the previous theme, provides cancer mortality statistics from 1953 to 1987 and allows maps and detailed tables to be tailor-made according to the interest of the user.

Methods

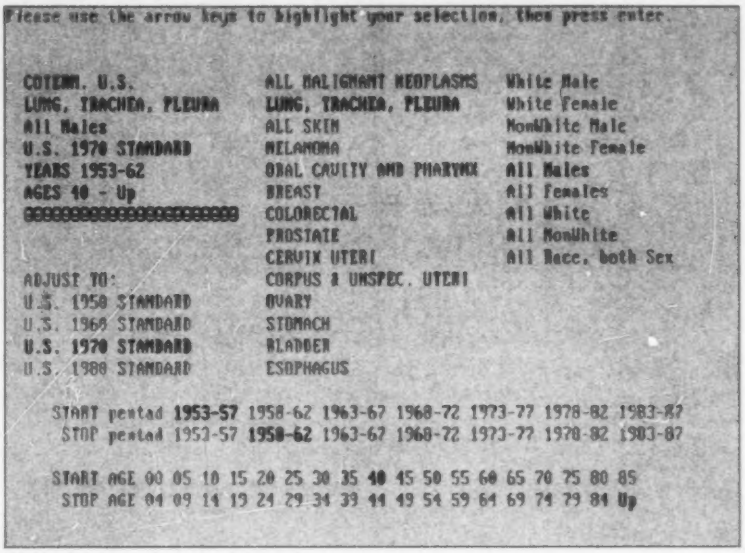
To encourage early cancer detection and prevention activities at the local level, the Map Program was developed by the National Cancer Institute with supplemental support from the Centers for Disease Control and Prevention and the American Cancer Society. This software program provides state and county mortality statistics in sufficient detail to allow physicians and health officials to analyze local patterns using mortality data from the National Center for Health Statistics and population data from the US Census Bureau.

After the user selects the state, type of cancer, and population parameters from the main menus, the county mortality rates are calculated, age adjusted, and tested for significance. Age-specific comparisons can also be made for each 5-year age group. Each county is color coded to show its mortality rate relative to the national rate for the same population parameters and period. Counties colored dark red have rates that are significantly higher than the national rate ($P < .05$) and 33% or more above it. Counties shown in light red have lower rates, but these rates are still significantly above the national average ($P < .05$). Yellow is for counties with rates similar to the national average, and green is for counties with rates significantly below the national average ($P < .05$). Once the map is created on the screen, a summary table can be printed that lists the counties

At the time of the study, Forrest A. Pommerenke was with the Early Detection Branch, National Cancer Institute, Bethesda, Md; he is now with the Department of Preventive Medicine, Medical College of Wisconsin, Milwaukee. Sudhir Srivastava is with the Early Detection Branch, and Robert W. Miller is with the Epidemiology Branch, National Cancer Institute. Susan P. Ackermann is with the Division of Cancer Prevention and Control, Centers for Disease Control and Prevention, Atlanta, Ga.

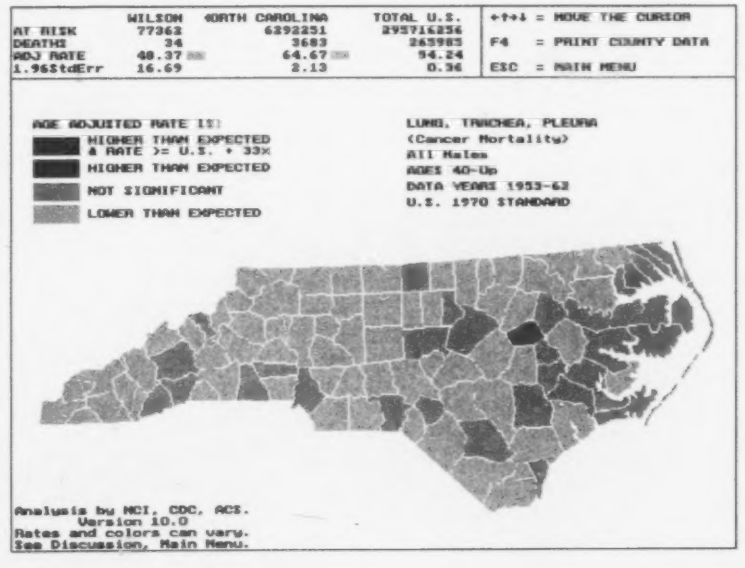
Requests for reprints should be sent to Forrest A. Pommerenke, MD, Department of Preventive Medicine, Medical College of Wisconsin, 8701 Watertown Plank Rd, Milwaukee, WI 53226.

This paper was accepted December 8, 1993.



Note. In this example, the parameters are entered for the map shown in Figure 2.

FIGURE 1—Main menu, State Cancer Control Map and Data Program.



Note. Cursor is in Wilson County.

FIGURE 2—Lung cancer mortality in North Carolina in all males aged 40 and older, 1953 through 1962.

according to the rates or number of deaths.

The user can then move the computer cursor to highlight a specific county on the screen. A display above the state map will appear, showing the number of deaths in the county, state, and nation, along with their age-adjusted rates and 95% confidence intervals. Once the cursor is within a county, a summary table can be printed that provides the number of deaths and age-specific rates for each age group for each 5-year period and the age-adjusted rates for the county.

Results

In the example in Figure 1, the highlighted items create a map showing the county mortality rates in North Carolina for all males 40 or more years of age who died of lung cancer from 1953 through 1962. The county rates were relatively low during this period; thus, the map appears predominantly green on the computer screen and light gray in Figure 2. This is in contrast to the much higher rates shown in the map for a more recent period (1978 through 1987), which appears predominantly yellow and red on the screen and as the medium to darker grays in Figure 3. In both figures the cursor has been placed in Wilson County, the data for which are shown at the top of the screen.

Encouraging Smoking Cessation

The county tables (not shown) printed from the Map Program summarize state and county data and can reveal trends in mortality. For example, the county summary table created for Wilson County, North Carolina (described earlier), shows a rise in lung cancer mortality from an age-adjusted rate of 48 per 100 000 in 1953 through 1962 to 283 per 100 000 in 1978 through 1987: a sixfold increase. The printed tables show that the greatest concentration of deaths is in men 55–74 years of age, but there is substantial mortality in younger men. The rate in the most recent interval was well above the US national rate of 199.73 during the same period (table, top of Figure 3).

The maps presented in Figures 2 and 3 could be used by public health officials to emphasize the importance of smoking cessation. Statistics taken from the Map Program tables showing ever-increasing mortality from lung cancer could provide a stimulus to local officials to restrict tobacco use in public places and to control youth access to tobacco.[10]

Increasing Cervical Screening of High-Risk Women

Figure 4 shows several Florida counties from 1968 through 1987, with cervical cancer mortality rates in non-White women that are three to five times higher

than those for White women. If the rates for non-Whites were reduced to those for Whites, the number of deaths due to cervical cancer among non-Whites in Duval–Jacksonville County, for example, would be reduced from 169 (table, top of Figure 4) to about 65. Data such as these can help focus attention on high-risk women who could benefit from increased cervical cancer screening.[11]

Implementing New Screening Procedures

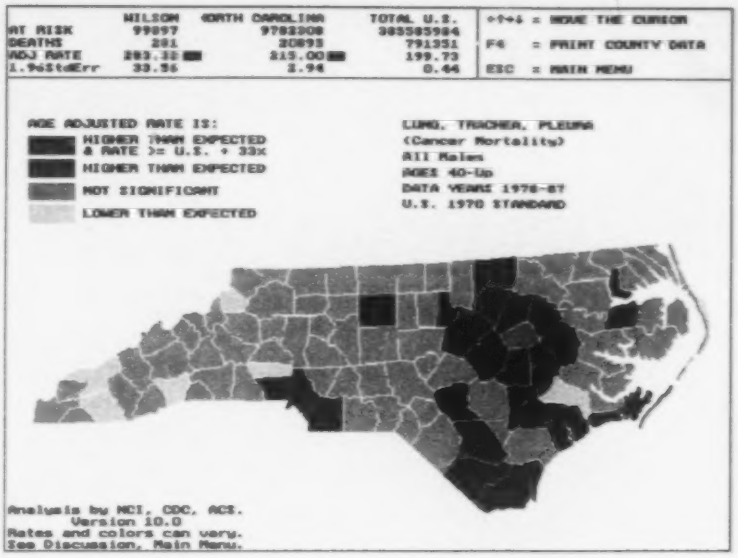
The *ras* oncogene, recently found in stool specimens from patients with curable colon cancer,[12] is one of a number of promising biomarkers that may find use in early detection.[13] When these markers are ready to be tested or used in screening high-risk groups, the Map Program can help determine where they might best be used: i.e., in a high-mortality-rate county among persons with the highest age-, sex-, and race-specific mortality rates. In regard to colorectal cancer, the Map Program reveals that one such location is Lackawanna County, Pennsylvania, where the colorectal mortality rate for all persons 40 or more years of age (83 per 100 000) was about 45% above the national average (57 per 100 000) from 1978 through 1987 (Figure 5).

Conclusions

The Map Program is easily used on personal computers. The versatility of the program is exemplified here by maps created according to parameters set by the user to show cancer mortality rates for several sites, by age group, for either sex, for Whites and/or non-Whites, and for different time intervals. Data from the program may also be used in correlation studies depending on the availability of other county-specific data bases. The program will be updated periodically.

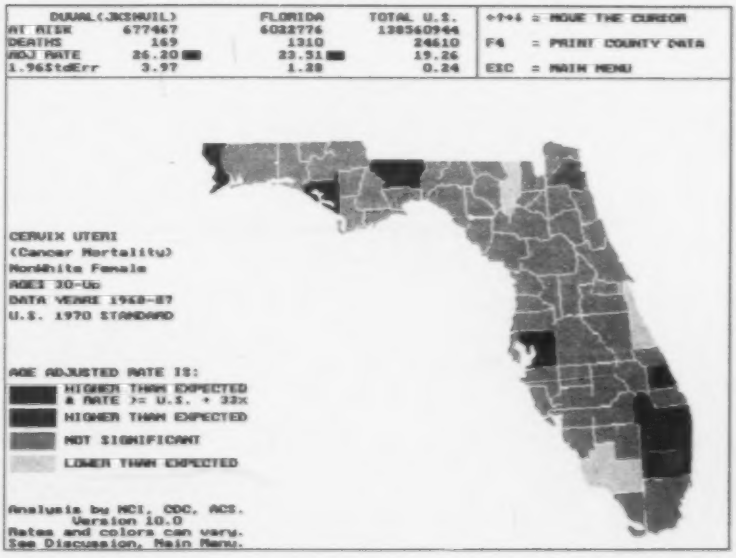
Public health officials should be cautioned that rates may be higher or lower by chance, especially when the population at risk is very small. The data responsible for abnormal rates should be reviewed carefully before resources are committed to extensive evaluations or interventions.[14,15]

Persons living in areas with high mortality rates can be particularly advised to take advantage of effective early detection procedures and preventive measures. In the future, populations at high risk for certain cancers may benefit from the use of biomarkers for early detection, chemopreventive agents, and diets, all of which are under investigation. □



Note. Cursor is in Wilson County. Note sharp contrast with figure 2.

FIGURE 3—Lung cancer mortality in North Carolina in all males aged 40 and older, 1978 through 1987.

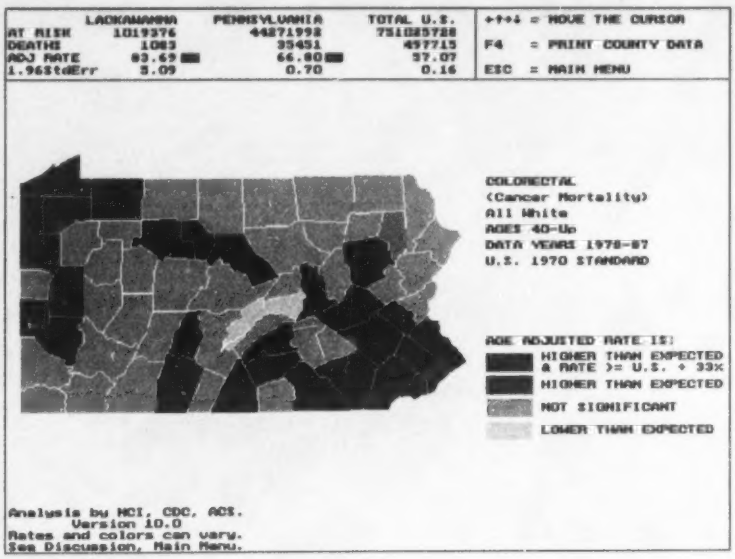


Note. The cursor is in Duval County (Jacksonville).

FIGURE 4—Cervical cancer mortality among non-White women in Florida, aged 30 and older, 1968 through 1987.

Acknowledgments

The State Cancer Control Map and Data Program was developed by the National Cancer Institute with support and technical assistance from the Centers for Disease Control and Prevention and the American Cancer Society. Programming was by Frank W. McKay.



Note. The cursor is in Lackawanna County.

FIGURE 5—Colorectal cancer mortality among Whites in Pennsylvania, aged 40 and older, 1978 through 1987.

The Map Program is available from the National Technical Information Service, Springfield, VA 22161, (703) 487-4650. The program requires a VGA board and monitor and DOS version 3.1 or higher.

References

1. National Cancer Institute. *Working Guidelines for Early Cancer Detection: Rationale and Supporting Evidence To Decrease Mortality.* Bethesda, Md: National Cancer Institute; 1987.
2. Shopland DR, Eyre JH, Pechacek TF. Smoking-attributable cancer mortality in 1991: is lung cancer now the leading cause of death among smokers in the United States? *J Natl Cancer Inst.* 1991;83:1143–1148.
3. Schottenfeld D, Fraumeni JS Jr, eds. *Cancer Epidemiology and Prevention.* Philadelphia, PA: WB Saunders Co; 1982.
4. Boss L, Suarez L. Data use for planning cancer prevention and control programs. *Public Health Rep.* 1990;105:354–360.
5. Mason TJ, McKay FW, Hoover R, Blot WJ, Fraumeni JF Jr. *Atlas of Cancer Mortality for U.S. Countries: 1950–1969.* Bethesda, Md: National Institutes of Health; 1975. DHEW publication NIH 75-780.
6. Fraumeni JF Jr. Etiologic insights from cancer mapping. *Princess Takamatsu Symposia.* 1988;18:13–25.
7. Mason TJ, McKay FW, Hoover R, Blot WJ, Fraumeni JF Jr. *Atlas of Cancer Mortality Among U.S. Nonwhites: 1950–1969.* Bethesda, Md: National Institutes of Health; 1976. DHEW publication NIH 76-1204.
8. Pickle LS, Mason TJ, Howard N, Hoover R, Fraumeni JF Jr. *Atlas of Cancer Mortality Among U.S. Nonwhites: 1950–1980.* Bethesda, Md: National Institutes of Health; 1990. DHHS publication NIH 90-1582.
9. Pickle LS, Mason TJ, Howard N, Hoover R, Fraumeni JF Jr. *Atlas of Cancer Mortality Among U.S. Whites: 1950–1980.* Bethesda, Md: National Institutes of Health; 1987. DHHS publication NIH 87-2900.
10. National Cancer Institute. *Strategies to Control Tobacco Use in the United States: A Blueprint for Public Health Action in the 1990's.* Bethesda, Md: National Institutes of Health; 1991. DHHS publication NIH 92-3316.
11. Hayward RA, Shapiro MF, Freeman HE, et al. Who gets screened for cervical and breast cancer? Results from a new national survey. *Arch Intern Med.* 1988;148:1177–1181.
12. Sidransky D, Tokino T, Hamilton S, Kinsler K, Levin B, Frost P, Vogelstein B. Identification of *ras* oncogene mutations in the stool of patients with curable colorectal tumors. *Science.* 1992;256:102–105.
13. Srivastava S, Kramer BS. Oncogenes in cancer detection. *Contemp Oncol.* 1992;2:63–72.
14. Neutra RR, Swan S, Freedman D, Mack T. Clusters galore. *Arch Environ Health.* 1990;45:314. Abstract.
15. Greenwald P, Nixon DW, Malone WF, Kelloff GJ, Stern HR, Witkin KM. Concepts in cancer chemoprevention research. *Cancer.* 1990;65:1483–1490.

Elevated Blood Lead Levels among Construction Workers in the Massachusetts Occupational Lead Registry

Richard Rabin, MSPH, Daniel R. Brooks, MPH, and Letitia K. Davis, ScD

ABSTRACT

Although the construction industry until recently was exempt from the Occupational Health and Safety Administration General Industry Lead Standard, including its medical monitoring provisions, periodic blood lead tests have been required for residential "deleaders" and structural painters in Massachusetts. Sixty-three percent of the 381 registrants in the Massachusetts Occupational Lead Registry with blood lead levels of 1.93 μmol/L or higher are construction workers. This proportion is much higher than that reported by registries of several states selected for comparison. These data highlight the need for better protection from lead exposure and the effectiveness of mandatory medical surveillance in identifying elevated blood lead levels among construction workers. (*Am J Public Health.* 1994;84:1483–1485)

Introduction

Published reports of overexposure to lead among construction workers are numerous. These reports come from two sources: case studies of particular work sites or projects[1–4] and presentations of aggregate data from state occupational lead registries.[5–8] In the registry reports, construction workers constitute only a small proportion of total registrants, although they comprise a larger percentage of those with the highest blood lead levels. This low rate of detection[9] is thought to be due to a lack of regular blood lead monitoring, rather than a low incidence of elevated blood lead levels. On the federal level, the construction industry has been exempt, until recently, from the Occupational Safety and Health Administration's (OSHA) Lead Standard,[10] which has medical monitoring requirements, and there are few states with regulations requiring blood lead monitoring of construction workers. An interim OSHA lead standard for construction workers was promulgated in May 1993.

The Massachusetts Occupational Lead Registry was established in April 1991. Since that time it has been collecting information from clinical laboratories on elevated blood lead levels among Massachusetts workers. The regulatory environment for construction workers in Massachusetts is different from that in most other states, and this difference has an impact on the number of reports received of elevated blood lead levels among construction workers.

Under Massachusetts workplace regulations, a subgroup of construction workers—residential lead abatement workers ("deleaders") and workers who remove lead paint from structures—are required to have periodic blood lead tests. Deleaders must be tested every 2 to 3 months, and all results must be reported by laboratories analyzing blood lead to the Massachusetts Department of Labor and Industries.[11] The regulations requiring blood lead monitoring among structural painters are much less precise. Workers need only be tested at the start of a project and then "periodically."[12] In practice, these regulations have not been enforced for house painters and enforced only sporadically for bridge painters (personal knowledge of one of the authors [R. R.] who works in the Massachusetts Department of Labor and Industries).

Despite these limitations, the existence of regulations mandating blood lead monitoring in Massachusetts presents an important opportunity to examine lead exposures among the construction work force. In this study, we review data on blood lead levels among construction workers reported to the Massachusetts Lead Registry during its first 2 years of operation. We also compare findings from Massachusetts with those from other states.

Methods

Clinical laboratories in Massachusetts must report the results of all blood lead tests over 0.72 μmol/L (15 μg/dL) for adults aged 15 years or older to the Department of Labor and Industries, Division of Occupational Hygiene.[13] Individuals with blood lead levels of 1.93 μmol/L (40 μg/dL) or greater are contacted. The follow-up procedure includes a telephone interview in which questions are asked concerning source of exposure and participation in a blood lead testing program. The present analysis is based on data reported to the Registry between May 1991 and April 1993. Data are based

Richard Rabin is with the Division of Occupational Hygiene, Massachusetts Department of Labor and Industries, West Newton, Mass. Daniel R. Brooks and Letitia K. Davis are with the Occupational Health Surveillance Program, Massachusetts Department of Public Health, Boston.

Requests for reprints should be sent to Richard Rabin, MSPH, Division of Occupational Hygiene, Massachusetts Department of Labor and Industries, 1001 Watertown St, West Newton, MA 02165.

This paper was accepted January 6, 1994.

TABLE 1—Construction Workers as a Proportion of Total Registrants with Occupational Exposure in the Massachusetts Lead Registry and in Lead Registries in Selected States

	Massachusetts	California	Texas	New Jersey	New York	Maryland
Years of data reviewed[a]	2	3¾	1	3½	2	3
Total number of registrants with occupational exposure[b]	311	1503	177	1870	3071	592
Percentage of registrants in construction industry	63.0	2.1	5.6	4.8	4.3	48.6
Employed males in construction industry—annual average	79 443	613 701	368 256	175 281	341 906	148 968
Annual rate of registrants per 100 000 in construction industry	123.4	1.3	2.7	14.5	19.2	64.4

Note. Data in the table are based on blood lead levels of greater than or equal to 1.93 μmol/L for Massachusetts, California, and Texas; and greater than or equal to 1.21 μmol/L for New Jersey, New York, and Maryland.

[a]Refers to years of data available for analysis. Registries may have been in existence for longer periods of time. Data are from the following time periods: Massachusetts, May 1991 through April 1993; California, April 1987 through December 1990; Texas, January through December 1989; New Jersey, January 1986 through June 1989; New York, January 1987 through December 1988; Maryland, January 1988 through December 1990.

[b]Based on cases with industry information available.

on the highest blood lead test for each registrant.

For this analysis, registrants were categorized as deleaders, other construction workers (primarily house and bridge [structural] painters), and nonconstruction workers. These categories are based on literal descriptions of employers obtained from physicians or through the registrant interviews.

The proportion of persons in the Massachusetts Lead Registry who are construction workers was compared with the proportions in five other states' registries. The proportion of employed Massachusetts construction workers appearing as registrants in the lead registry was also compared with the proportions in those states.

Data for comparison with the Massachusetts Registry were available from the California, Maryland, New Jersey, New York, and Texas registries. Maryland has been one of the very few states with a comprehensive standard for exposure to lead among construction workers, including a requirement for blood lead monitoring.[14] The remaining states, including Massachusetts, have not had such a standard.

Comparisons were based on the most recent data available in published form from each state that specifically identified construction workers.[7,15–18] The blood lead level for which industry information could be determined varied by state. Data are presented for blood lead levels beginning at 1.93 μmol/L for Massachusetts, California, and Texas and at 1.21 μmol/L (25 μg/dL) for Maryland, New Jersey, and New York. We calculated the number of registrants in the construction industry per 100 000 employed male construction workers from employment figures from the Bureau of Labor Statistics' Current Population Survey using the appropriate years for each state.[19] Rates were calculated based on employment of male construction workers, because virtually all construction industry registrants were male.

Results

Of 381 individuals with blood lead levels of 1.93 μmol/L or greater reported to the Massachusetts Lead Registry, 311 (82%) were determined to have resulted from occupational exposure. Construction workers constituted 63% (n = 196) of this group. Deleaders comprised 60% (n = 118) of construction workers with blood leads exceeding 1.93 μmol/L; bridge and house painters accounted for almost all the rest (36%, n = 70).

Of 49 workers with blood lead levels of 2.90 μmol/L (60 μg/dL) or greater, 39 (80%) worked in construction. Twenty (51%) were bridge and house painters, and seventeen (44%) were deleaders.

The Massachusetts Registry was able to contact and interview 177 workers (57%). All interviewed deleaders (n = 51) reported that their blood test was part of an employer-sponsored program, compared with only 37% (n = 19) of other construction workers and 78% (n = 50) of nonconstruction workers.

Construction workers comprise a much larger proportion of registrants with occupational exposures in Massachusetts and Maryland (63% and 49%, respectively) than in California, New Jersey, New York, and Texas, where the proportions range from 2% to 6% (Table 1). The difference in the number of registrants in construction per 100 000 construction workers in each state is even greater. The Massachusetts rate of 123.4 was about twice that of Maryland (64.4) and 6 to 100 times higher than that of the other states.

Discussion

The finding that construction workers constitute 63% of occupationally exposed individuals with blood lead levels greater than 1.93 μmol/L in the Massachusetts Lead Registry sharply contrasts with reports from most other states. There are at least two likely explanations. First, the greater detection of elevated blood lead levels in Massachusetts construction workers is likely attributable to the mandatory blood lead testing of deleaders and structural painters described above. Second, the total number of deleaders, including those with elevated blood lead, may be greater, given the extensive, legally mandated deleading activity in the state. Residential lead paint abatement is required in Massachusetts whenever a child under the age of 6 lives in a dwelling unit with lead paint, a requirement virtually unique in the nation.[20] As a result, although compliance is not complete, a large number of residential units are deleaded each year, more than 4500 in 1991.[21]

Alternatively, the high proportion of construction workers in the Massachusetts Registry might be attributed to a possible relative absence of workers in

other lead-using industries. However, the rate of elevated blood lead levels of construction workers in the Massachusetts Lead Registry far exceeds the rates for other states, which suggests that increased detection and frequency of elevated blood lead levels are more likely explanations. Even the Massachusetts rate should be considered an underestimate, given the presumed low rate of monitoring among large groups of construction workers exposed to lead.

The large number of deleaders in the Registry is cause for concern. It underscores the hazardous nature of deleading and the need for both increased enforcement of deleading regulations and further research to evaluate alternative lead removal and lead control measures. Some states have banned dry scraping,[22,23] which continues to be a common form of lead removal in Massachusetts. The extent to which elevated blood lead levels in this industry can be attributed to the lead removal process or inadequate control techniques remains to be further evaluated.

It is notable that the majority of construction workers with blood lead levels greater than 2.90 μmol/L are not deleaders but workers in the structural painting industry—namely, bridge painters and house painters. Some of these workers may face higher levels of exposure or have less adequate controls than deleaders. In addition, they are much less likely than deleaders to participate in routine blood lead monitoring. Structural painters may be more likely to receive testing when their blood leads reach higher levels (i.e., when they are more likely to be symptomatic). Some states, including Massachusetts, have begun incorporating language specifying comprehensive lead protection measures in all public contracts for paint removal on bridges.

The Massachusetts deleading law appears to be more effective in generating blood lead monitoring among deleaders compared not only with structural painting workers, but also with nonconstruction workers, although the latter are covered by the OSHA lead standard, which requires blood lead monitoring. The better compliance in the deleading industry is likely due to the fact that the law requires not only monitoring of employees but also the reporting of results, and such reporting is a requirement for continued licensure. In developing new laws and regulations, officials should consider placing an affirmative obligation on employers to report blood lead test results to state authorities as an added measure for ensuring compliance.

The data from the Massachusetts Occupational Lead Registry, in conjunction with findings from lead registries in other states, indicate that hazardous lead exposure among construction workers is almost certainly grossly underreported nationally. The extent of overexposure to lead among construction workers demonstrated by the Massachusetts Lead Registry should not be surprising, given the history of lead paint use in this country. Concerns about protection for construction workers have been raised since long before 1978, when the OSHA Lead Standard for General Industry went into effect. This study supports the need for strict enforcement of the recently promulgated OSHA Lead Standard for Construction, which provides workers in this industry with new protections. □

Acknowledgments

This work was supported in part by a grant (U60/CCU108415-01) from the National Institute for Occupational Safety and Health.

We wish to thank Clelia Alburez for her excellent work in data collection and management.

References

1. Centers for Disease Control. Lead poisoning in bridge demolition workers—Massachusetts. *MMWR.* 1989;38:687–694.
2. Centers for Disease Control. Lead exposures among lead burners—Utah, 1991. *MMWR.* 1992;41:307–310.
3. Pollock CA, Ibels LS. Lead intoxication in Sydney Harbour Bridge workers. *Aust N Z J Med.* 1988;18:46–52.
4. Marino PE, Franzblau A, Lilis R, Landrigan PJ. Acute lead poisoning in construction workers: the failure of current protective standards. *Arch Environ Health.* May/June 1989;44:140–145.
5. Centers for Disease Control. Surveillance for occupational lead exposure—United States, 1987. *MMWR.* 1989;38:642–646.
6. Waller K, Osorio AM, Maizlish N, Royce S. Lead exposure in the construction industry: results from the California Occupational Lead Registry, 1987 through 1989. *Am J Public Health.* 1992;82:1665–1671.
7. Tepper A. Surveillance of occupational lead exposure in New Jersey: 1986 to 1989. *Am J Public Health.* 1992;82:275–277.
8. Baser ME, Marion D. A statewide case registry for surveillance of occupational heavy metals absorption. *Am J Public Health.* 1990;80:162–164.
9. Rudolph L, Sharp DS, Samuels S, Perkins C, Rosenberg J. Environmental and biological monitoring for lead exposure in California workplaces. *Am J Public Health.* 1990;80:921–925.
10. Code of Federal Regulations: Lead. Washington, DC: Occupational Safety and Health Administration; 1978. 29 CFR §1910.1025.
11. Code of Massachusetts Regulations: Deleading. Boston, Mass: Department of Labor & Industries; 1988. 454 CMR §22.00.
12. Code of Massachusetts Regulations: Structural Painting. Boston, Mass: Department of Labor & Industries; 1981. 454 CMR §11.00.
13. Code of Massachusetts Regulations: Occupational Lead Poisoning Registry. Boston, Mass: Department of Labor & Industries; 1991. 454 CMR §23.00.
14. Code of Maryland Regulations: Occupational Exposure to Lead in Construction Work. Baltimore, Md: Department of Licensing and Regulation; 1984. COMAR §09.12.32.
15. Maizlish N, Rudolph L. California adults with elevated blood lead levels, 1987 through 1990. *Am J Public Health.* 1993;83:402–405.
16. *Elevated Blood Lead Levels in Maryland Workers.* Baltimore, Md: Center for Environmental Health, Maryland Department of the Environment; annual reports 1988–1990.
17. *Annual Summaries of the Heavy Metals Registry.* Albany, NY: Division of Occupational and Environmental Health, New York State Department of Health; 1987–1988.
18. *Epidemiology in Texas, 1989.* Austin, Tex: Epidemiology Division, Texas Department of Health; 1989.
19. *Geographic Profile of Employment and Unemployment.* Washington, DC: Bureau of Labor Statistics; 1987–1991 editions: Table 17.
20. Code of Massachusetts Regulations: Lead Poisoning Prevention and Control. Boston, Mass: Department of Public Health; 1990. 105 CMR §460.00.
21. Office of Technical Services. Unpublished data. Boston, Mass: Massachusetts Department of Labor and Industries; data provided in 1993.
22. Code of Maryland Regulations. Baltimore, Md: Department of the Environment; 1988. COMAR §26.02.06
23. Code of Connecticut Regulations. Hartford: Department of Health Services; 1992. §19a-111.

The Adequacy of Prenatal Care Utilization Index: Its US Distribution and Association with Low Birthweight

Milton Kotelchuck, PhD, MPH

ABSTRACT

The proposed Adequacy of Prenatal Care Utilization Index is applied to data from the 1980 National Natality Survey to assess the adequacy of prenatal care utilization and its association with low birthweight in the United States. The index suggests that only 61.1% of women received adequate prenatal care, including 17.7% with more intensive care; 16.7% received inadequate care. More White women (63.4%) than Black women (51.9%) received adequate prenatal care. Low-birthweight rates were elevated among women with inadequate prenatal care and among those who received more intensive prenatal care. (*Am J Public Health.* 1994;84:1486–1489)

Introduction

The measurement of prenatal care utilization adequacy depends on the accuracy of the index used; previous efforts may have overestimated the adequacy of prenatal care utilization in the United States by incorrectly assessing utilization after enrollment. Elsewhere in this issue of the Journal, I discuss limitations of the Kessner/Institute of Medicine Index[1] and propose an alternative two-part Adequacy of Prenatal Care Utilization (APNCU) Index.[2]

In this paper I apply the APNCU Index to data from the 1980 National Natality Survey, a representative sample of US births. Characteristics of the database have been described previously.[3,4] The goals of this paper are to assess the adequacy of prenatal care utilization in the United States and its first-order association with low birthweight.

Distribution of Adequacy of Prenatal Care Utilization in the United States

The last column of Table 1 shows the distribution of adequacy of prenatal care utilization in the United States as judged by the two dimensions of the APNCU Index and by the summary (combined) index. Overall, 61.1% of US women received adequate or better prenatal care, including 17.7% with more intensive Adequate Plus care; 38.9% received less than adequate prenatal care, including 16.7% with Inadequate care. (The criteria used for these designations are described in Kotelchuck.[2])

"Adequacy of Initiation of Prenatal Care," or month prenatal care begins, shows the well-known prenatal care initiation pattern in the United States: 87.5% of women initiated care before month 5, and only 4.6% began after month 6 or not at all.

"Adequacy of Received Services," or percentage of recommended visits received (the dimension of prenatal care newly described by the APNCU Index), suggests a more mixed picture of prenatal care utilization. On a positive note, 68.5% of women received adequate or better prenatal care after initiation of care. Indeed, 22.2% received additional prenatal care visits beyond the number recommended by the American College of Obstetricians and Gynecologists.[5] Unfortunately, 31.4% of women received less than optimal care, with 6.7% receiving fewer than 50% of the recommended visits.

Table 1 also shows the adequacy of prenatal care utilization stratified by race. Black women had substantially poorer overall prenatal care utilization; almost twice as many Black women (27.8%) as White women (14.0%) received Inadequate care. The most profound racial difference is the delayed initiation of prenatal care within the Black community; 21.1% of Black women, compared with 10.5% of White women, began care after month 4. After initiation of prenatal care, there is less disparity but more variability in utilization; Black women were more likely than Whites both to receive too few visits and to receive more visits than the recommended number.

Table 2 shows the full distribution of the two dimensions of the APNCU Index. Women whose care was rated Inadequate (16.7%) generally did not receive Inadequate ratings on both dimensions; only 2.5% were rated Inadequate on both. Women with no care (1.0%; data not shown) were not the major contributors to the Inadequate ratings for either dimension.

The author is with the Department of Maternal and Child Health, School of Public Health, University of North Carolina at Chapel Hill.

Requests for reprints should be sent to Milton Kotelchuck, PhD, MPH, Department of Maternal and Child Health, School of Public Health, University of North Carolina at Chapel Hill, CB# 7400, Chapel Hill, NC 27599-7400.

This paper was accepted September 23, 1993.

Editor's Note. See related editoral by Wise (p 1374) and article by Kotelchuck (p 1414) in this issue.

Association of Prenatal Care Utilization with Low Birthweight

Table 3 examines the association of the two dimensions of the APNCU Index with low birthweight. (These are first-order associations; they are not meant to prove or disprove the impact of increased prenatal care utilization on improved birth outcomes. Such an analysis would require a much more sophisticated methodology than is possible in this paper.) There is a positive linear association between month of initiation of prenatal care and low-birthweight rates. Low-birthweight rates increased substantially for women who initiated prenatal care after the 4th month.

There appears to be a U-shaped association between recommended visits received and low-birthweight rates; both women with Inadequate ratings and those with Adequate Plus ratings had increased low-birthweight rates. The best birth outcomes were seen in women who started prenatal care in the first half of pregnancy and received between 50% and 109% of the recommended number of visits. Those women who were rated lowest on both dimensions had the worst birth outcomes (18.9% had low-birthweight babies), although this result might have been due to confounding with other concurrent social risks and premature delivery before enrollment in prenatal care. Women with no prenatal care had a low-birthweight rate of 26.7% (data not shown).

Table 4 reflects both the distribution of prenatal care utilization (Table 2) and the rate of low birthweight associated with that distribution (Table 3). It demonstrates the distinct distributions of adequacy of prenatal care utilization among all low-birthweight infants and among total births in the United States. The majority of low-birthweight births (63.0%) occurred to women with Adequate or Adequate Plus prenatal care. Women who received extra prenatal services (17.7%) had a disproportionate share of low-birthweight births (38.5%; crude relative risk, 2.17). Women with Intermediate or Adequate prenatal care had disproportionately fewer low-birthweight infants. Women with Inadequate care (16.7%) also had elevated proportions of low-birthweight births (24.5%). The 2.5% of women whose care was rated Inadequate on both dimensions had the most elevated relative risk, but these women accounted for only 5.3% of all low-birthweight

TABLE 1—Ratings Assigned to Births According to the Adequacy of Prenatal Care Utilization Index, by Maternal Race

	White, %	Black, %	All Births, %
Month of initiation of care			
7–9 or no care	4.0	7.0	4.6
5–6	6.5	14.1	7.8
3–4	34.3	40.5	35.4
1–2	55.1	38.4	52.1
% of recommended visits received			
0–49	5.5	11.6	6.7
50–79	24.6	25.0	24.7
80–109	48.3	37.5	46.3
110+	21.6	25.8	22.2
Summary index			
Inadequate	14.0	27.8	16.7
Intermediate	22.6	20.3	22.2
Adequate	45.8	32.6	43.4
Adequate Plus	17.6	19.3	17.7

Note. See Kotelchuck[2] for explanation of the Adequacy of Prenatal Care Utilization Index. Percentage may not add to 100 because of rounding.
Source. Percentages are based on an analysis of the birth certificate data from the 1980 National Natality Survey.[3]

TABLE 2—Percentage Distribution of US Births Rated on the Two Dimensions of the Adequacy of Prenatal Care Utilization Index

	% of Recommended Visits Received				
Month of Initiation of Care	0–49	50–79	80–109	≥110	% of All Births
7–9 or no care	1.6	0.5	0.9	1.6	4.6
5–6	0.9	2.0	2.1	2.9	7.8
3–4	1.9	9.6	16.3	7.6	35.4
1–2	2.3	12.6	27.0	10.1	52.1
% of all births	6.7	24.7	46.3	22.2	100.0

Note. See Kotelchuck[2] for explanation of the Adequacy of Prenatal Care Utilization Index. Percentages may not add to 100 because of rounding.
Source. Percentages are based on an analysis of the birth certificate data from the 1980 National Natality Survey.[3]

TABLE 3—Low-Birthweight Percentages for Women Whose Prenatal Care Was Rated on the Two Dimensions of the Adequacy of Prenatal Care Utilization Index

	% of Recommended Visits Received				
Month of Initiation of Care	0–49	50–79	80–109	≥110	All Births
7–9 or no care	18.9	10.4	8.0	7.9	11.9
5–6	7.5	7.7	7.3	13.4	9.7
3–4	9.0	4.6	3.9	15.7	6.9
1–2	9.1	3.3	3.9	14.5	6.1
All births	11.3	4.3	4.1	14.3	6.9

Note. See Kotelchuck[2] for explanation of the Adequacy of Prenatal Care Utilization Index. Percentages may not add to 100 because of rounding.
Source. Percentages are based on an analysis of the birth certificate data from the 1980 National Natality Survey.[3]

TABLE 4—Distributions of Total and All Low-Birthweight Births, by Adequacy of Prenatal Care Utilization (APNCU) Index

APNCU Category	Total Births, %	All Low-Birthweight Births, %	Ratio
Inadequate	16.7	24.5	1.47
Intermediate	22.2	12.5	.56
Adequate	43.4	24.5	.56
Adequate Plus	17.7	38.5	2.17

Note. See Kotelchuck[2] for explanation of the Adequacy of Prenatal Care Utilization Index.
Source. Percentages are based on an analysis of the birth certificate data from the 1980 National Natality Survey.[3]

infants in the United States (data not shown).

Discussion

The APNCU Index suggests a more somber picture of prenatal care utilization in the United States than do the other two major prenatal care indexes. The APNCU Index suggests that 61.1% of women in the United States receive adequate prenatal care, compared with the 65.9% suggested by the Kessner Index and the 77.9%[2] suggested by the trimester of prenatal care initiation. Similarly, the APNCU Index suggests that more women have inadequate care (16.7%) than is suggested by either the Kessner Index (7.7%) or the trimester of prenatal care initiation (4.6%). The lower rates of prenatal care utilization adequacy indicated by the APNCU Index reflect the index's more complete assessment and the added contribution of prenatal care utilization after initiation. Overall, the APNCU Index ratings imply that one in six American women have inadequate prenatal care and almost two in five have less than adequate care. These are truly disgraceful figures for our country.

Both the timing of initiation and the number of prenatal care visits contribute to the poor overall prenatal care utilization rates. The fact that care rated Inadequate on one of the APNCU Index's two dimensions is not generally rated Inadequate on the other suggests that recruiting women into the prenatal health care system and keeping them there may be two distinct phenomena. The APNCU Index suggests that late initiation of care is the more significant problem in the United States, which agrees with recent data from cross-national comparisons of prenatal care utilization.[6]

The poorer prenatal care utilization by Black women than by White women suggested by the APNCU Index is consistent with the results of other national analyses.[7–9] Delayed initiation of prenatal care, not received services, appears to be the principal reason for the racial disparity in utilization. That more Black women than White women receive Adequate Plus services can be interpreted positively as an indication of the health care system's responsiveness to higher-risk Black women; unfortunately, however, more Black women are also not receiving minimal prenatal care services.

The U-shaped association of low birthweight and prenatal care utilization indicated by the APNCU Index is clinically plausible and has been noted previously.[10] Extensive prenatal care visits have been linked to high-risk pregnancies.[11,12] But perhaps because the predominant Kessner Index has no classification for intensive prenatal care services, the association between low birthweight and prenatal care has been predominantly described in a more linear fashion, emphasizing poor birth outcomes directly related to poor prenatal care. Women who receive extra prenatal care visits are clearly a very high-risk group, but they are already known to health care providers. The efficacy of prenatal care interventions, rather than access, may be the important issue for these women. Their high proportion of low-birthweight births should make us somewhat sanguine about proposals to improve US infant mortality rates that simply recommend more prenatal care visits generally or for high-risk women only.

The relationship between prenatal care utilization and low birthweight is determined in part by the measure of prenatal care utilization. Given the elevated low-birthweight rates among women receiving Adequate Plus services, the weak correlations between the Kessner Index and birth outcomes reported in the public health literature may now be better appreciated.[11,13] The majority of low-birthweight births in the United States occur to women with adequate or better prenatal care utilization, irrespective of whether one uses the Kessner Index (52.2%), trimester of prenatal care initiation (70.1%), or the APNCU Index (63.0%) (data not shown). An Inadequate rating on the APNCU Index is, however, associated with a higher percentage of the total low-birthweight distribution (24.5%) than an Inadequate rating on the Kessner Index (14.9%) or the trimester of prenatal care initiation (8.0%). The ability of the APNCU Index to more accurately and comprehensively assess prenatal care utilization should enhance our understanding of the association between prenatal care utilization and birth outcomes in the United States. □

Acknowledgments

An earlier version of this paper was presented at the annual meeting of the American Public Health Association, New Orleans, La, October 18–21, 1987.

The author wishes to thank Paul Wise, Chris Costello, Diane Ramsey, and Chris Little for their help in the computation and interpretation of the APNCU Index data presented in this paper.

References

1. Kessner DM, Singer J, Kalk CE, Schlesinger ER. *Infant Death: An Analysis by Maternal Risk and Health Care.* Washington, DC: Institute of Medicine and National Academy of Sciences; 1973:chap 2.
2. Kotelchuck M. An evaluation of the Kessner Adequacy of Prenatal Care Index and a proposed Adequacy of Prenatal Care Utilization Index. *Am J Public Health.* 1994;84: 1414–1420.
3. Placek PJ. The 1980 National Natality and Fetal Mortality Surveys. *Public Health Rep.* 1984;99:111–116.
4. Keppel KG, Heuser RL, Placek PJ. Methods and response characteristics, 1980 National Natality and Fetal Mortality Surveys. *Vital Health Stat* [2]. 1986; no. 100. DHHS publication PHS 86-1374.
5. *Standards for Obstetric-Gynecologic Services.* 6th ed. Washington, DC: American College of Obstetricians and Gynecologists; 1985.
6. Buekens P, Kotelchuck M, Blondel B, et al. A comparison of prenatal care use in the United States and Europe. *Am J Public Health.* 1993;83:31–36.
7. Kotelchuck M, Costello CA, Wise PH. Race differences in utilization of prenatal care services in the U.S. Presented at the 115th Annual Meeting of the American Public Health Association; October 18–21, 1987; New Orleans, La.
8. Virji SK, Cottington E. Risk factors associated with preterm deliveries among racial groups in a national sample of married mothers. *Am J Perinatol.* 1991;8:347–353.
9. National Center for Health Statistics. Ad-

vance report of final natality statistics, 1989. *Monthly Vital Stat Rep.* December 12, 1991;40(8)(suppl 1).
10. Harris JE. Prenatal medical care and infant mortality. In: Fuchs VR, ed. *Economic Aspects of Health.* Chicago, Ill: University of Chicago Press; 1982:15–52.
11. Institute of Medicine. *Preventing Low Birthweight.* Washington, DC: National Academy Press; 1985.
12. Gordis L, Kleinman JC, Klerman LV, Mullen PD, Paneth N. Criteria for evaluating evidence regarding the effectiveness of interventions. In: Merkatz IR, Thompson JE, eds. *New Perspectives on Prenatal Care.* New York, NY: Elsevier Science Publishing Co; 1990:31–38.
13. Liberatos P, Kiely JL. Selected issues in the evaluation of prenatal care. In: Kiely M, ed. *Reproductive and Perinatal Epidemiology.* Boca Raton, Fla: CRC Press, 1991:79–98.

The Effect of Paternal Smoking on the Birthweight of Newborns Whose Mothers Did Not Smoke

Fernando D. Martinez, MD, Anne L. Wright, PhD, Lynn M. Taussig, MD, and the Group Health Medical Associates

ABSTRACT

We examined the relationship of paternal smoking habits to birthweight in a cohort of infants enrolled at birth (n = 1219). To assess validity of parental smoking reports, cord serum levels of cotinine were measured in 175 newborns. A mean birthweight deficit of 88 g was found in newborns of nonsmoking mothers whose fathers smoked more than 20 cigarettes/day. Only 2 of 138 newborns whose mothers said they were nonsmokers had cotinine levels indicating that their mothers were active smokers. Among infants of nonsmoking mothers, detectability of cotinine in cord blood was significantly correlated with the number of cigarettes smoked daily by fathers. These data suggest that fetal growth may be adversely affected when the mother is passively exposed to tobacco smoke during pregnancy. (*Am J Public Health.* 1994;84:1489–1491)

Introduction

There is strong evidence indicating that active maternal smoking during pregnancy is associated with reduced birthweight in the mothers' offspring.[1,2] Deficits ranging from 90 to 300 g have been reported. However, available data on a possible effect of mothers' passive exposure to cigarette smoke on birthweight are less convincing. MacMahon et al.[3] reported that infants whose fathers smoked had a mean birthweight 85 g less than that of infants whose fathers did not smoke. Martin and Bracken[4] reported that passive smoking by nonsmoking mothers was significantly related to delivering a low-birthweight newborn, after adjustment for confounding factors. Rubin et al.[5] found that fathers' smoking was associated with a mean reduction in birthweight of 6 g per cigarette smoked. It has been suggested, however, that a significant proportion of these effects were attributable to smoking mothers who were classified as nonsmokers based on their responses to questionnaires.[6,7]

The aim of this study was to determine the possible effect of paternal smoking on birthweight. To assess the influence of misclassification as a possible source of bias, levels of cotinine (a metabolite of nicotine) were measured in cord blood of a representative sample of the population.

Subjects and Methods

The study subjects were enrollees of the Tucson Children's Respiratory Study, a longitudinal study of risk factors for lower respiratory tract illnesses in infancy and childhood.[8] Newborns whose parents used one of the largest health maintenance organizations in Tucson, Arizona, were enrolled. Premature infants were not excluded if they were healthy. A total of 1596 infants were eligible, but participation was refused for 350 infants (22%). There were no statistically significant differences in terms of parental smoking, parity, maternal employment, or household size between families who refused to participate and those who enrolled their newborns into the study.[8] However, enrolled and refusal groups differed significantly in terms of marital status, parental age, parental education, and ethnicity.[8]

Information on birthweight, maternal and paternal education, number of previous births, duration of pregnancy, and ethnicity was obtained by study nurses while mother and newborn were still in the hospital. Each parent was given a questionnaire concerning, among other issues, his or her smoking habits. Specifically, mothers and fathers were asked whether they smoked regularly and the number of cigarettes smoked daily (none, 1–10, 11–20, and > 20). Because questionnaires were answered within the first month of life of the child (median = 15 days), parents who said they smoked were

The authors are with the Department of Pediatrics and the Respiratory Sciences Center, University of Arizona College of Medicine, Tucson, Ariz. For a list of the Group Health Medical Associates, see the acknowledgments.

Requests for reprints should be sent to Fernando D. Martinez, MD, Department of Pediatrics, University of Arizona Health Sciences Center, 1501 N Campbell Ave, Tucson, AZ 85724.

This paper was accepted March 30, 1994.

TABLE 1—Infants' Birthweight, by Maternal and Paternal Smoking

Smoking (Cigarettes/Day)	No. Subjects	Birthweight,[a] g (Mean ± SD)
Maternal[b]		
None	992	3596 ± 401
1–10	43	3424 ± 358
11–20	95	3374 ± 445
>20	38	3323 ± 396
R^2		.037
P for linearity		<.0001
Paternal, with nonsmoking mother[c]		
None	716	3602 ± 401
1–10	63	3573 ± 388
11–20	68	3520 ± 382
>20	60	3514 ± 428
R^2		.005
P for linearity		.026

[a]Adjusted for gestational age, gender, ethnicity, birth order, maternal and paternal education, and maternal age.
[b]Information on birthweight was not available for three infants; data on maternal smoking were unavailable for one infant. The number of cigarettes smoked per day was not available for 41 smoking mothers. Information about at least one confounder was unavailable for 33 subjects.
[c]Newborns whose mothers were single, separated, or divorced (n = 67) were not included. The number of cigarettes smoked per day was not reported for 18 smoking fathers.

TABLE 2—Multiple Regression Analysis for Children of Nonsmoking Mothers with Birthweight as the Dependent Variable (n = 907)

	Difference in Birthweight, g	95% CI
Gestational age	−166	−219, −113
Hispanic	−55	−135, 25
Other non-Anglo	−104	−257, 49
Birth order	78	45, 111
Paternal education	8	−70, 86
Maternal education	−52	−126, 22
Maternal age	6	0, 12
Paternal smoking	−34	−63, −5
Constant[a]	3607	...

Note. Coding for variables was gestational age in weeks: 40 weeks = 0; gender: males = 0, females = 1; Hispanic: no = 0, yes = 1; other non-Anglo: no = 0, yes = 1; birth order: first = 0, second = 1, third = 2, fourth + = 3; paternal and maternal education: >12 years = 0, ≤12 years = 1; maternal age in years: 30 years = 0; paternal smoking: nonsmoker = 0, 1–10 cigarettes/day = 1, 11–20 cigarettes/day = 2, >20 cigarettes/day = 3. CI = confidence interval.
[a]The constant represents the expected birthweight for an Anglo boy, first born to nonsmoking parents with college-level education; gestational age = 40 weeks and maternal age = 30 years. Coefficients indicate change in birthweight (in grams) when compared with the reference category in each group (with a value of 0).

considered to have smoked during pregnancy.

Cord blood specimens were collected for 87% of the enrolled infants; cord serum cotinine levels were measured by competitive radioimmunoassay[9] in a sample of 175 infants (14%). There was no significant difference in any social or demographic parameter between these 175 infants and the rest of the eligible population. The lower limit of detectability for the serum cotinine assay was 1 ng/mL.

Statistical Analyses

Birthweights were adjusted for possible confounders by standard regression procedures. The Mantel test was used to test for linear trends in proportions.[10] This procedure was chosen because it is particularly suitable for sparse data.[11] Logistic regression was used in the subgroup of infants who had cotinine measured to determine whether maternal and paternal smoking independently increased the proportion of infants with detectable cotinine levels.

Results

Relationship between Paternal Smoking and Birthweight

Duration of pregnancy was not significantly affected by either maternal or paternal smoking. There was a significant correlation between birthweight and the number of cigarettes smoked by the mother (Table 1). Newborns of mothers who reported smoking more than 20 cigarettes per day had a mean deficit of 273 g. When newborns of nonsmoking mothers were studied separately (Table 1), a significant trend for birthweight to decrease with increasing numbers of cigarettes smoked by the father was observed. Newborns of nonsmoking mothers whose fathers smoked more than 20 cigarettes per day had a mean deficit of 88 g. This association was independent of maternal and paternal education, birth rank, ethnicity, maternal age, duration of pregnancy, and sex. The independent contribution of these variables to birthweight among children of nonsmoking mothers is shown in Table 2.

Validation of Questionnaire Data and Assessment of Misclassification

To assess the validity of questionnaire data, cotinine levels were measured in cord serum of a representative sample of 175 newborns. The mothers of 37 of these subjects reported being smokers. Only 7 of these 37 infants (19%) had undetectable cotinine in cord blood. All 23 subjects whose mothers smoked more than 10 cigarettes per day had detectable cotinine levels, compared with half of those whose mothers smoked 1–10 cigarettes per day and 4.4% of those whose mothers were nonsmokers ($P < .0001$). Median (range) cotinine concentrations were 3.5 ng/mL (0–99 ng/mL) for newborns whose mothers smoked 1–10 cigarettes per day; 59 ng/mL (2–195 ng/mL) for newborns whose mothers smoked 11–20 cigarettes per day; and 55 ng/mL (14–239 ng/mL) for newborns of mothers who smoked more than 20 cigarettes per day. It is evident from the above data that no newborns of mothers who smoked 11 or more cigarettes per day had serum cotinine levels of <2 ng/mL. Moreover, all seven newborns of mothers who reported smoking 1–10 cigarettes per day and who had detectable serum cotinine had ≥2 ng/mL.

Cotinine levels were undetectable in 130 of 138 (94%) newborns whose mothers reported being nonsmokers, and an additional 6 subjects (4.3%) had serum cotinine levels at the lower limit of detectability for this assay (1 ng/mL). The remaining two subjects (1.5%) had cord serum cotinine concentrations of 23 and 101 ng/mL, respectively. These values were clearly within ranges found only among newborns of active smokers. Because values of 1 ng/mL were not seen among infants whose mothers reported being smokers, we estimated that in our population misclassification of smoking mothers as nonsmokers was approximately 1.5% (95% confidence interval = 0.5, 3.5).

We studied the relation between paternal smoking and detectability of cotinine in 130 infants of nonsmoking mothers (after excluding the 2 misclassified infants and those of an additional six mothers who were single, separated, or divorced). In this group, cord serum cotinine was detectable in only 2.7% (3 of 110) of the newborns with nonsmoking fathers, whereas cotinine was detectable in 15.0% (3 of 20) of the newborns whose mothers were nonsmokers but whose fathers smoked ($P < .05$). The frequency of detectable cotinine in cord blood was also strongly correlated with the number of cigarettes smoked by the father ($P = .001$). A logistic regression was performed with detectability of cord serum cotinine as the dependent variable and smoking by each parent (in four levels, as described previously) as the independent variable. Both maternal smoking ($P < .0001$) and paternal smoking ($P = .01$) were independent determinants of the presence of detectable cord serum cotinine levels.

Discussion

In this study we demonstrated a significant influence of paternal smoking on birthweight in newborns of nonsmoking mothers, and this effect was independent of several possible confounders. Our results are consistent with those of Haddow and collaborators,[12] who reported a deficit of 107 g in newborns of nonsmoking mothers who had serum cotinine levels of 1.1–9.9 ng/mL during pregnancy. Our findings also confirm the report by Rubin et al.[5] of a significant reduction in birthweight associated with paternal smoking. After controlling for confounders, they found that an infant born to a father who smoked 20 cigarettes per day had an average reduction in birthweight of 120 g. This compares well with our value of 88 g for an infant born to a father who smoked more than 20 cigarettes per day. Recent reports from Italy[13] and India[14] also suggest significant effects of paternal smoking on birthweight.

It has been suggested that the association between passive maternal smoking and birthweight is explained by underreporting of active maternal smoking.[6,7] To assess this possibility, we measured cotinine levels in cord blood in a representative sample of the study subjects. Only two infants (1.5%) of mothers who reported being nonsmokers had cotinine levels within the range of those found in newborns of smoking mothers. In both cases, the father was a nonsmoker. Thus, it is improbable that the small degree of underreporting of active smoking by mothers may account for our findings.

We cannot completely exclude the possibility that differences in life-style or nutritional status between persons exposed and unexposed to environmental tobacco smoke may explain our findings.[15] However, reports of lower birthweight in newborns whose mothers were exposed to environmental tobacco smoke during pregnancy come from countries with very different life-styles[5,12–14]; this fact argues against the possibility that a specific confounder or combination thereof may explain the consistent effects observed in these studies. We did not ask mothers about alcohol consumption during pregnancy, but Rubin et al.[5] and Haddow et al.[12] showed that this factor does not explain the relationship between the mother's exposure to environmental tobacco smoke and birthweight. Finally, other sources of exposure not considered in this study, such as presence of other smokers at home or in the workplace, seemed to be considerably less important than paternal smoking; only 2.7% of newborns of nonsmoking parents had detectable cord serum cotinine levels.

The clinical implications of reductions in birthweight such as those reported herein are not known. Birthweight is inversely correlated with both neonatal and postneonatal mortality.[16] The increased risk is not limited to low-birthweight infants, however; infants with birthweights of 3000–3499 g have infant mortality rates that are higher than those of infants weighing 3500–3999 g at birth.[16] Thus, it is conceivable that the reduction in birthweight caused by both active and passive maternal smoking may increase the risk for mortality during the first year of life.

Our study adds new evidence to the increasing body of information showing unfavorable effects of passive smoking on human health. Pregnant women should not only be advised to stop smoking; they should also be encouraged to avoid environments in which cigarettes are smoked by others. □

Acknowledgments

This research was supported by a Specialized Center of Research grant (HL14136) from the National Institutes of Health.

The Group Health Medical Associates staff are Drs John Bean, Henry Bianci, John Curtiss, John Ey, Robert Moss, James Rothschild, Alejandro Sanguineti, Barbara Smith, Terry Vondrak, and Neil West; and Maureen McLellan, RN, PNP.

References

1. Rush D, Cassano P. Relationship of cigarette smoking and social class to birth weight and perinatal mortality among all births in Britain. *J Epidemiol Community Health.* 1983;37:249–255.
2. Comstock GW, Shah FK, Meyer MB, Abbey H. Low birth weight and neonatal mortality rate related to maternal smoking and socioeconomic status. *Am J Obstet Gynecol.* 1971;111:53–59.
3. MacMahon B, Alpert M, Salber EJ. Infant weight and parental smoking habits. *Am J Epidemiol.* 1966;82:247–261.
4. Martin TR, Bracken MB. Association of low birth weight with passive smoke exposure in pregnancy. *Am J Epidemiol.* 1986; 124:633–642.
5. Rubin DH, Krasilnikoff PA, Leventhal JM, Weile B, Berget A. Effect of passive smoking on birth-weight. *Lancet.* 1986;2: 415–417.
6. Trichopoulos D. Passive smoking, birthweight, and oestrogens. *Lancet.* 1986;2:743.
7. MacArthur C, Knox EG. Passive smoking and birthweight. *Lancet.* 1987;1:37–38.
8. Taussig LM, Wright AL, Morgan WJ, Harrison MR, Ray CG, GHMA Pediatricians. The Tucson Children's Respiratory Study, I: design and implementation of a prospective study of acute and chronic respiratory illness in children. *Am J Epidemiol.* 1989;129:1219–1231.
9. Langone JJ, Gjika HB, Van Vunakis H. Nicotine and its metabolites: radioimmunoassays for nicotine and cotinine. *Biochemistry.* 1973;12:5025–5030.
10. Mantel N. Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J Am Stat Assoc.* 1963;58:690–700.
11. Rothman KJ. *Modern Epidemiology.* 1st ed. Boston, Mass: Little, Brown; 1986:346.
12. Haddow JE, Knight GJ, Palomaki GE, McCarthy JE. Second-trimester serum cotinine levels in nonsmokers in relation to birth weight. *Am J Obstet Gynecol.* 1988;159: 481–484.
13. Lazzaroni F, Bonassi S, Manniello E, et al. Effect of passive smoking during pregnancy on selected perinatal parameters. *Int J Epidemiol.* 1990;19:960–966. 14.
Mathai M, Vijayasri R, Babu S, Jeyaseelan L. Passive smoking and birth weight in a south India population. *Br J Obstet Gynaecol.* 1992;99: 342–343. 15.
Koo LC, Ho JH, Rylander R. Life-history correlates of environmental tobacco smoke: a study of nonsmoking Hong Kong Chinese wives with smoking and nonsmoking husbands. *Soc Sci Med.* 1988;26:751–760. 16.
Hogue CJR, Buehler JW, Strauss LT, Smith JC. Overview of the National Infant Mortality Surveillance (NIMS) project—design, methods, results. *Public Health Rep.* 1987;102:126–138.

Intrauterine Cocaine, Lead, and Nicotine Exposure and Fetal Growth

Daniel R. Neuspiel, MD, MPH, Morri Markowitz, MD, and Ernest Drucker, PhD

ABSTRACT

Intrauterine tobacco, lead, and cocaine exposure often co-occur and may affect fetal growth and development, yet studies of gestational cocaine effects have not adequately measured lead or tobacco. In this anonymous survey, blood lead and urine cotinine levels were determined and mothers were queried about tobacco use. Eighteen cocaine-exposed mother–infant dyads had higher lead and cotinine levels than 46 random nonexposed dyads, regardless of reported cigarette smoking. Crude growth decrements in cocaine-exposed newborns were attenuated after control for lead and cotinine. Future studies of gestational cocaine effects should measure other toxic exposures with more precision. (*Am J Public Health* 1994;84:1492–1495)

Introduction

Gestational cocaine exposure has been associated with fetal growth[1] and infant behavioral deficits,[2] yet possible confounding from other risk factors[3,4] has not been fully explored. Tobacco and lead have also been associated with impaired fetal growth and brain development. Cocaine users may be exposed to lead in older substandard housing associated with poverty and may have both active and passive tobacco exposure. Further, tobacco and lead may interact: active smokers[5–7] and their children[8–11] have higher lead levels, with a dose–response gradient between parental smoking and child lead levels[12,13]; newborn lead levels have been correlated with the duration of maternal smoking.[14] This investigation hypothesized that cocaine-exposed newborns have more lead and tobacco exposure and that cocaine-associated effects on the growth of newborns are partly explained by lead and/or tobacco exposure.

Methods

Over 2 months in 1992, parturient women in a public hospital provided urine samples for routine clinical testing, the discarded portions of which were separated in two for drug and cotinine measurement. At delivery, umbilical cord blood was placed in a heparinized lead-free tube for lead assay. After study staff obtained clinical information from hospital records, the study urine and blood samples were given a number code and other identifiers were removed. Urine samples were tested in women proceeding to singleton deliveries. When more than one urine sample was provided, the earliest sample was used.

One urine sample was assayed for cocaine, benzoylecgonine, opiates, methadone, cannabinoids, and amphetamine by the EMIT method.[15] Two randomly selected comparison dyads with negative toxicology were chosen for each drug-positive case. All women giving birth were routinely asked about cocaine and tobacco use during pregnancy by clinical staff, independent of this study. The other study urine sample was assayed for cotinine (a metabolite of nicotine) by radioimmunoassay, and creatinine was determined to correct for variation in urine volume.[16]

Whole blood lead was measured by atomic absorption spectrophotometry in a laboratory participating in proficiency testing programs of the Centers for Disease Control and New York State. The 95% confidence limit for blood lead is ± 1 μg/dL.[17]

Statistical analysis was done with SYSTAT software.[18] This investigation was approved by the Institutional Review Board of Montefiore Medical Center and by the New York City Health and Hospitals Corporation. Because study data were collected anonymously, informed consent was waived.

Results

Discard urine specimens were obtained from 237 (41%) of the 581 mothers having singleton deliveries during the study period. Cocaine-positive cases were identified from study assays (10 of 237 tested), clinical tests (7 additional cases of 47 tested), or maternal history (1 additional case). Five case subjects with study urine samples positive only for drugs other than cocaine were excluded. Forty-six comparison subjects were chosen from 220 dyads with both negative study toxicology and no self-report of drug use. Of 18 cocaine-positive dyads, 6 were tested for both cotinine and lead, 6 only for cotinine, and 6 only for lead. Of 46 cocaine-negative dyads, 20 had both lead and cotinine levels, 13 had only cotinine, and

Daniel R. Neuspiel is with the Department of Pediatrics and the Department of Epidemiology and Social Medicine, Morri Markowitz is with the Department of Pediatrics, and Ernest Drucker is with the Department of Epidemiology and Social Medicine, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY.

Requests for reprints should be sent to Daniel R. Neuspiel, MD, MPH, NCBH 4M-08, 3424 Kossuth Ave, Bronx, NY 10467.

This paper was accepted February 23, 1994.

12 had only lead. There were no differences between the subjects with and without anonymous urine samples either in the proportion of infants clinically tested for drugs (6% and 9%, respectively) or in the proportion of clinical tests that were positive for cocaine (29% and 32%, respectively); the chi-square test was nonsignificant for each.

Demographic, pregnancy, and newborn information is in Table 1. Of 64 mothers, 56% were Hispanic, 31% African-American, 5% White, 6% Asian, and 2% other ethnicity. The mean lag time from urine collection to testing was similar in both exposure groups.

Lead and cotinine levels (Table 2) were higher in the cocaine group, in which cotinine levels were elevated both among cigarette smokers and nonsmokers. Lead levels differed between admitted smokers (mean = 5.2 μg/dL; 95% confidence interval [CI] = 4.5, 5.9) and nonsmokers (mean = 3.5 μg/dL; 95% CI = 3.0, 4.0; $P = .001$).

Multiple linear regression (Table 3) was used to evaluate the independent effects of cocaine, lead, and cotinine on fetal growth after control for gestational age. The apparent effects of cocaine on birthweight and head circumference were diminished after consideration of cotinine levels; smoking history explained fewer cocaine effects than cotinine and was a weaker confounder. Lead exposure explained some of the cocaine-related decrement in all three growth measures, but it had less effect than cotinine on weight and head circumference. Separate models demonstrated no confounding of cocaine effects on fetal growth by race, maternal age, gravidity, or parity. Nine regressions were conducted with each of the three dependent variables.

TABLE 1—Characteristics of Mothers and Newborns, by Cocaine Exposure

	Group				
	Cocaine Exposed (n = 18)		Comparison (n = 46)		
	Mean	95% CI	Mean	95% CI	*P*
Maternal age, y	30.5	28.6, 32.4	23.8	22.3, 25.3	<.0005
Gravidity	5.6	4.2, 7.0	2.7	2.2, 3.2	.001
Previous live births	1.8	1.2, 2.4	0.9	0.6, 1.2	.012
Involuntary pregnancy terminations	0.9	0.4, 1.5	0.3	0.1, 0.5	.034
Elective pregnancy terminations	1.8	1.1, 2.5	0.5	0.3, 0.7	.004
Maternal race, % Black	61	38, 84	20	8, 32	.001
Smoking report this pregnancy, %	56	33, 79	13	3, 23	<.0005
Female births, %	44	21, 67	43	29, 57	.944
Gestational age, wk	37.6	36.1, 39.1	39.1	38.7, 39.6	.071
Birthweight, g	2953	2767, 3139	3364	3227, 3501	.001
Length, cm	47.9	46.7, 49.1	50.9	50.3, 51.5	<.0005
Head circumference, cm	32.4	31.6, 33.3	33.8	33.4, 34.3	.009

TABLE 2—Lead and Cotinine Levels of Mother–Infant Dyads, by Cocaine Exposure

	Group						
	Cocaine Exposed			Comparison			
	No.	Mean	95% CI	No.	Mean	95% CI	*P*
Lead (μg/dL)	12	4.75	3.98, 5.52	32	3.38	2.80, 3.95	.01
Cotinine[a] (ng per mg of creatinine)	12	1176	325, 4260	33	46	24, 89	<.0005
Among smokers	6	3395	1444, 7979	5	1486	328, 6730	.38
Among nonsmokers	6	407	45, 3673	28	25	16, 38	.06

[a]Because cotinine levels were skewed, log transformation was done before statistical calculations; geometric means are reported.

Discussion

Many methodologic difficulties exist in studying cocaine effects on the fetus.[2] The effects of concurrent exposures are of particular importance because cocaine users share many other potential perinatal risk factors.

Cocaine-exposed newborns may be exposed to more ambient lead. Tobacco exposure may also raise lead levels by small-airway damage and more hand-to-mouth behavior.[13] Nutritional factors may predispose cigarette smokers and cocaine users to higher lead levels via inadequate iron or calcium intake. In addition, crack smoking increases lung permeability due to chronic epithelial damage,[19] which may increase absorption of airborne lead. Maternal alcohol intake has been related to lead levels, independent of smoking.[9] Also, behavioral manifestations of chronic lead exposure may lead to more cocaine use.

Lead and cocaine have a demonstrated neurotoxicological interaction in animals.[20–22] Intrauterine lead[23] and tobacco[24] have been associated with later cognitive and behavioral problems. Consideration of these exposures as confounders is therefore important in studying the long-term neurodevelopmental effects of intrauterine cocaine.

Tobacco affects the fetus,[25] and cocaine-using women actively smoke more cigarettes[3,4] and may be exposed to more passive tobacco smoke. The reported number of cigarettes smoked per day weakly correlates with urine cotinine levels.[26] These data show that reported maternal smoking inadequately controls for the effects of tobacco on fetal growth compared with quantitative urine cotinine levels.

Incomplete cohort participation in the current study was primarily due to lack of 24-hour supervision of specimen collection. Cocaine users may have been reluctant to provide urine specimens because of fearing punitive consequences, even though they were unaware of the study. However, nonparticipants had similar rates of clinical toxicology testing. The negative-toxicology group may have included some clandestine cocaine users, because the urine tests in this study were only sensitive for use within 3 days, but then the differences detected may underestimate the truth.

The groups differed in several health and social factors that could affect growth

TABLE 3—Correlates of Cocaine Exposure and Fetal Growth: Multiple Linear Regression

Dependent Variable (y)[a]	Additional Independent Variables[a]	Regression Coefficient[b]	P[c]
Birthweight, g	x_3 = None	−311	.02
	x_3 = Cotinine	−144	.42
	x_3 = Smoking history	−204	.15
	x_3 = Lead	−239	.23
	x_3, x_4 = Cotinine, lead	−54	.87
Head circumference, cm	x_3 = None	−1.08	.02
	x_3 = Cotinine	−0.75	.15
	x_3 = Smoking history	−0.96	.07
	x_3 = Lead	−0.87	.15
	x_3, x_4 = Cotinine, lead	−0.18	.78
Length, cm	x_3 = None	−2.78	<.0005
	x_3 = Cotinine	−2.84	<.0005
	x_3 = Smoking history	−2.90	<.0005
	x_3 = Lead	−2.32	.03
	x_3, x_4 = Cotinine, lead	−1.86	.08

[a]Separate regression equations (using model $y = x_0 + x_1 + x_2 \ldots x_n$) each include a constant (x_0), cocaine exposure (x_1), and gestational age (x_2) as independent variables.
[b]Change in dependent variable associated with cocaine exposure, controlling for other independent variables in model.
[c]P value for effect of cocaine exposure.

outcomes. Although separate regressions showed no effects of these factors, there was insufficient power to control for them simultaneously or to examine for statistical interaction. There was no information on important factors such as alcohol, cocaine dose, or exposure period; and the dichotomization of cocaine exposure may have diminished the true effects. It was not apparent why birth length was less influenced by cotinine than were the other growth parameters.

Many factors associated with cocaine use in the environments of pregnant women and young children may affect development and behavior. Blaming cocaine alone for adverse outcomes may have negative maternal consequences, such as unnecessary termination of pregnancy.[27] The labeling of cocaine-exposed children as irreparably damaged "crack kids" may confer a self-fulfilling prophecy of failure and ignore other remediable causes of developmental impairment.[28,29] Cocaine users and their children have many other risks that need more accurate consideration. A one-dimensional emphasis on cocaine is an inaccurate causal explanation, leading to inappropriate public health policies. □

Acknowledgments

This research was supported by Grant 1-R18-DA06925 from the National Institute on Drug Abuse.

This paper was presented in part at the 55th Annual Scientific Meeting of the College on Problems of Drug Dependence, Toronto, Canada, June 16, 1993.

Camille Hall and Hadassa Padilla assisted in data collection. The American Health Foundation performed urine cotinine and creatinine assays, and lead assays were done by the Division of Pediatric Metabolism of Montefiore Medical Center.

References

1. Zuckerman B, Frank DA, Hingson R, et al. Effects of maternal marijuana and cocaine use on fetal growth. *N Engl J Med.* 1989;320:762–768.
2. Neuspiel DR, Hamel SC. Cocaine and infant behavior. *J Dev Behav Pediatr.* 1991;12:55–64.
3. Neuspiel DR, Hamel SC, Hochberg E, Greene J, Campbell D. Maternal cocaine use and infant behavior. *Neurotoxicol Teratol.* 1991;13:229–233.
4. Frank DA, Zuckerman BS, Amaro H, et al. Cocaine use during pregnancy: prevalence and correlates. *Pediatrics.* 1988;82:888–895.
5. Baghurst PA, McMichael AJ, Vimpani GV, Robertson EF, Clark PD, Wigg NR. Determinants of blood lead concentrations of pregnant women living in Port Pirie and surrounding areas. *Med J Aust.* 1987;146:69–73.
6. Hense HW, Filipiak B, Novak L, Stoeppler M. Nonoccupational determinants of blood lead concentrations in a general population. *Int J Epidemiol.* 1992;21:753–762.
7. Hopper JL, Balderas A, Mathews JD. Analysis of variation in blood lead levels in Melbourne families. *Med J Aust.* 1982;2:573–576.
8. Baghurst PA, Tong S, McMichael AJ, Robertson EF, Wigg NR, Vimpani GV. Determinants of blood lead concentrations to age 5 years in a birth cohort study of children living in the lead smelting city of Port Pirie and surrounding areas. *Arch Environ Health.* 1992;47:203–210.
9. Ernhart CB, Wolf AW, Sokol RJ, Brittenham GM, Erhard P. Fetal lead exposure: antenatal factors. *Environ Res.* 1985;38:54–66.
10. McMichael AJ, Baghurst PA, Vimpani GV, Robertson EF, Wigg NR, Tong SL. Sociodemographic factors modifying the effect of environmental lead on neuropsychological development in early childhood. *Neurotoxicol Teratol.* 1992;14:321–327.
11. Rothenberg SJ, Schnaas L, Cansino-Ortiz S, et al. Neurobehavioral deficits after low level lead exposure in neonates: the Mexico City pilot study. *Neurotoxicol Teratol.* 1989;11:85–93.
12. Andren P, Schutz A, Vahter M, et al. Environmental exposure to lead and arsenic among children living near a glassworks. *Sci Total Environ.* 1988;77:25–34.
13. Willers S, Schutz A, Attewell R, Skerfving S. Relation between lead and cadmium in blood and the involuntary smoking of children. *Scand J Work Environ Health.* 1988;14:385–389.
14. Bellinger DC, Needleman HL, Leviton A, Waternaux C, Rabinowitz MB, Nichols ML. Early sensory-motor development and prenatal exposure to lead. *Neurobehav Toxicol Teratol.* 1984;6:387–402.
15. Jatlow P. Cocaine: analysis, pharmacokinetics, and metabolic disposition. *Yale J Biol Med.* 1988;61:105–113.
16. Langone J, Gjika HB, Van Vunakis H. Nicotine and its metabolites: radioimmunoassay for nicotine and cotinine. *Biochemistry.* 1973;12:5015–5025.
17. Markowitz M, Rosen JF. Need for the lead mobilization test in children with lead poisoning. *J Pediatr.* 1991;119:305–310.
18. SYSTAT for Windows, version 5 edition. Evanston, Ill: SYSTAT; 1992.
19. Susskind H, Weber DA, Volkow ND, Hitzemann R. Increased lung permeability following long-term use of free-base cocaine (crack). *Chest.* 1991;100:903–909.
20. Burright RG, Donovick PJ, Michels K, Fanelli R, Dolinsky Z. Effect of amphetamine and cocaine on seizure in lead treated mice. *Pharmacol Biochem Behav.* 1982;16:631–635.
21. Cory-Slechta DA, Widzowski DV. Low level lead exposure increases sensitivity to the stimulus properties of dopamine D_1 and D_2 agonists. *Brain Res.* 1991;553:65–74.
22. Grover CA, Nation JR, Bratton GR. Chronic exposure to lead attenuates cocaine-induced behavioral activation. *Pharmacol Biochem Behav.* 1992;44:221–225.
23. Needleman HL, Bellinger D. The health effects of low level exposure to lead. *Annu Rev Public Health.* 1991;12:111–140.
24. Rush D. Exposure to passive cigarette smoking and child development: an updated critical review. In: Poswillo D, Alberman E, eds. *Effects of Smoking on the Fetus, Neonate, and Child.* New York, NY: Oxford University Press; 1992:150–170.

25. Zhang J, Ratcliffe JM. Paternal smoking and birthweight in Shanghai. *Am J Public Health.* 1993;83:207–210.
26. Wall MA, Johnson J, Jacob P, Benowitz NL. Cotinine in the serum, saliva, and urine of nonsmokers, passive smokers, and active smokers. *Am J Public Health.* 1988;78: 699–701.
27. Koren G, Gladstone D, Robeson N, Robieux I. The perception of teratogenic risk of cocaine. *Teratology.* 1992;46:567–571.
28. Mayes LC, Granger RH, Bornstein MH, Zuckerman B. The problem of prenatal cocaine exposure: a rush to judgment. *JAMA.* 1992;267:406–408.
29. Neuspiel DR. On pejorative labeling of cocaine exposed children. *J Subst Abuse Treat.* 1993;10:407.

Selected Antepartum Medical Complications and Very-Low-Birthweight Infants among Black and White Women

Michael DeBaun, MD, MPH, Diane Rowley, MD, MPH, Michael Province, PhD, Joseph W. Stockbauer, MS, and F. Sessions Cole, MD

ABSTRACT

This study estimated the risk of very-low-birthweight delivery among Black and White women with selected treatable antepartum medical conditions. A logistic regression model was applied to a retrospective, population-based data set identified by computerized, linked birth certificate and maternal hospital discharge records. For Black mothers, the adjusted odds ratio for very-low-birthweight delivery was statistically significant for essential hypertension and urinary tract infection. For White mothers, the adjusted odds ratio was statistically significant for essential hypertension, urinary tract infection, pregnancy-induced hypertension, and diabetes mellitus. Public policy designed to reduce the risk of very-low-birthweight delivery must include strategies for attenuating the impact of treatable antepartum medical conditions. (*Am J Public Health.* 1994;84: 1495–1497)

Introduction

Recently, Kempe et al. suggested that 80% of the excess risk of very-low-birthweight (<1500 g) delivery among Black infants results from four maternal medical complications: chorioamnionitis or premature rupture of membranes, idiopathic preterm labor, hypertensive disorders, and hemorrhage.[1] With the exception of hypertensive disorders, none of the antepartum maternal medical conditions studied by Kempe et al. can be effectively treated to prevent preterm delivery. We investigated whether potentially treatable maternal medical conditions increase the risk of very-low-birthweight delivery. In the data available for this study, very low birthweight serves as the best available surrogate for preterm delivery.

Methods

The study population consisted of subsets of infants born to St. Louis County and Kansas City, Mo, residents during the period 1983 through 1985. We identified these infants from the linked birth certificate and maternal hospital discharge data set maintained by the Center for Health Statistics of the Missouri Department of Health. This linked data set provided more complete information on maternal medical conditions than birth certificates alone.[2,3]

The study was restricted to 272 Black and 292 White singleton, live-born, very-low-birthweight (<1500 g) infants and 9820 Black and 36 763 White infants of normal birthweight (>2499 g). We excluded infants with birthweights of 1500 to 2499 g to reduce the number of growth-retarded infants in the data set and sharpen the focus on preterm delivery.[4] Linked birth and maternal discharge records were available for 11 185 Black and 38 574 White births (70% and 80% of total births, respectively) (Table 1). For both races, the linked and nonlinked populations had similar demographic characteristics (Table 1).

All complications were recorded in the *International Classification of Diseases,* 9th Revision, Clinical Modification (ICD-9-CM) coding system. The complications were divided into the following categories: diabetes mellitus (ICD-9-CM codes 250 to 2500, 2502 to 2508, 6480), gestational diabetes (ICD-9-CM code 6488); pregnancy-induced hypertension (ICD-9-CM codes 6423 to 6429), essential hyperten-

Michael DeBaun and F. Sessions Cole are with the Edward Mallinckrodt Department of Pediatrics, Washington University School of Medicine, St. Louis Children's Hospital, St. Louis, Mo. Diane Rowley is with the National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Ga. Michael Province is with the Division of Biostatistics, Washington University. Joseph W. Stockbauer is with the Missouri Department of Health, Jefferson City.

Requests for reprints should be sent to F. Sessions Cole, MD, Division of Newborn Medicine, St. Louis Children's Hospital, One Children's Pl, St. Louis, MO 63110.

This paper was accepted December 28, 1993.

TABLE 1—Singleton Infant and Maternal Characteristics, by Race: St. Louis County and Kansas City, Mo, 1983 through 1985

	Birth and Maternal Discharge Records, %			
	Blacks		Whites	
	Linked (n = 11 185)	Unlinked (n = 4824)	Linked (n = 38 574)	Unlinked (n = 9910)
Infant birthweight, g				
>2499	87.8	86.4	95.3	93.5
<1500	2	3.1	0.8	1.3
Maternal characteristics				
Less than 18 years old	10.5	10.1	2.2	2.9
More than 35 years old	3.4	2.7	5.1	4.2
Less than high school education	68.2	71.9	48.4	49.4
Single vs married	61.6	65.7	9.3	13.4
Smoked during pregnancy	28.3	33.6	25.7	24.4
Inadequate prenatal care	53.8	50.1	15.0	21.6
Parity ≥3 vs <3	11.8	12.4	6.7	7.5
Body mass index[a] < 18	7.4	8.6	7.6	9.0

[a]Weight in kilograms divided by height in meters squared. The Centers for Disease Control and Prevention classifies body mass index less than 18 as extremely underweight.

TABLE 2—Crude and Adjusted Odds Ratios for Very-Low-Birthweight (<1500 g) Delivery, by Race: St. Louis County and Kansas City, Mo, 1983 through 1985

Maternal Medical Condition	Odds Ratio (95% Confidence Interval)			
	Blacks		Whites	
	Crude	Adjusted[a]	Crude	Adjusted[a]
Diabetes mellitus	0.8 (0.3, 2.7)	1.0 (0.3, 3.3)	2.5 (1.2, 5.4)	2.7 (1.2, 6.2)
Gestational diabetes	1.2 (0.3, 4.9)	1.6 (0.4, 6.5)	1.3 (0.3, 5.2)	1.5 (0.3, 6.3)
Essential hypertension	3.3 (1.3, 8.2)	4.9 (1.9, 12.8)	4.5 (1.4, 14.3)	5.2 (1.6, 17.0)
Pregnancy-induced hypertension	1.6 (1.0, 2.4)	1.6 (1.0, 2.6)	3.9 (2.7, 5.5)	4.2 (2.9, 6.0)
Urinary tract infection	3.1 (1.9, 5.0)	2.8 (1.6, 4.7)	6.4 (4.0, 10.5)	5.6 (3.3, 9.2)

[a]Adjusted for parity, body mass index, maternal age, marital status, cigarette smoking, educational attainment, and prenatal care.

sion (ICD-9-CM codes 4010 to 4019), and urinary tract infections (ICD-9-CM codes 590 to 5999, 595 to 5909, 599.0, 6466).

For each selected maternal medical condition, we estimated crude and adjusted odds ratios (with 95% confidence intervals) of the risk of having very-low-birthweight infants for Blacks and Whites separately because of the race-specific impact of sociodemographic factors. We used a logistic regression model to determine the adjusted odds ratio for each maternal medical complication and to determine the effect of race for each such complication after accounting for the effect of previously identified confounders of very-low-birthweight delivery. These confounders, obtained from the birth certificate, were parity (<3, ≥3), body mass index (weight in kilograms divided by height in meters squared) based on prepregnancy weight and height,[5] maternal age (<18 years, 18 through 34 years, >34 years), marital status (single, other), cigarette smoking during pregnancy (present, absent), educational attainment (completed <12 years of school, completed ≥12 years of school), and prenatal care (Kessner index: adequate, inadequate).[5]

Results

The prevalences of delivery of very-low-birthweight infants were 2.4% for Blacks and 0.8% for Whites (see Table 1). For normal-birthweight infants (>2499 g), Black mothers had higher rates (per 1000 live births) of the following antepartum medical conditions than did White mothers: gestational diabetes (6.9 vs 5.8), diabetes mellitus (13.0 vs 8.7), essential hypertension (10.3 vs 2.3), pregnancy-induced hypertension (124.1 vs 35.43), and urinary tract infection (65.5 vs 10.8).

Among Black mothers, both the crude and adjusted odds ratios and 95% confidence intervals for risk of very-low-birthweight delivery were greater than 1.0 for essential hypertension and urinary tract infection (see Table 2). Although neither the adjusted nor the crude odds ratio for pregnancy-induced hypertension was statistically significant for Black mothers, the magnitude of the point estimates suggests that the complication may be clinically relevant.

Among White mothers, the crude and adjusted odds ratios and 95% confidence intervals for very-low-birthweight delivery were greater than 1.0 for each condition except gestational diabetes. White mothers with pregnancy-induced hypertension had a greater risk of very-low-birthweight delivery than Black mothers ($P = .03$). There were no statistical differences between the risk of very-low-birthweight delivery in Blacks and Whites for gestational diabetes ($P = .90$), diabetes mellitus ($P = .30$), essential hypertension ($P = .31$), or urinary tract infection ($P = .24$).

Discussion

Maternal medical complications in these data contribute to the risk of very-low-birthweight delivery and vary in estimated effects among Black and White mothers. Essential hypertension among Black mothers and urinary tract infection among White mothers were associated with the greatest risk of very-low-birthweight delivery.

In a similar population-based study, Kempe et al.[1] also found a significantly increased risk of very-low-birthweight delivery associated with maternal conditions in data abstracted from medical records. The maternal conditions reported, however, typically result in delivery within hours or days after diagnosis. In contrast, we selected conditions that do not immediately precipitate delivery since they are more amenable to medical interventions.

Our study design had several limitations inherent in linked record data sets. Objective parameters were not used to define each maternal medical complica-

tion since all maternal conditions were identified from medical discharge diagnoses. Another limitation was the inability to assess the effect of medical interventions on each maternal condition. In addition, we were unable to account for conditions that occurred early in the pregnancy but resolved, since the maternal conditions were based on hospital maternal discharge diagnosis at the time of birth. Furthermore, we could not detect the effect of maternal medical conditions at different gestational ages, since gestational age was not consistently and accurately measured. Despite these limitations inherent in linked record data sets, previous studies with similar study designs have provided important findings regarding the effects of maternal smoking on infant mortality and preterm delivery.[6,7]

We conclude that treatable maternal antepartum medical complications contribute significantly to the risk of very-low-birthweight delivery among Black and White mothers but vary in estimated effects by race. Future research in maternal and child health is warranted to assess the impact of treating maternal conditions associated with very-low-birthweight delivery. □

Acknowledgments

We wish to thank Paul Wise, MD, MPH, for his thoughtful suggestions and Amy Papian for excellent secretarial support.

References

1. Kempe A, Wise PH, Barkan SE, et al. Clinical determinants of the racial disparity in very low birth weight. *N Engl J Med.* 1992;327:969–973.
2. Carucci PM. *Reliability of Statistical and Medical Information Reported on Birth and Death Certificates.* Albany, NY: New York State Department of Health; 1979. New York Department of Health monograph 15.
3. Van Metre PR. *An Accuracy Assessment of the Medical Supplement to the Utah Live Birth Certificate.* Salt Lake City, Utah: University of Utah; 1977. Thesis.
4. Villar J, Belizan JM. The relative contribution of prematurity and fetal growth retardation to low birth weight in developing and developed societies. *Am J Obstet Gynecol.* 1982;143:793–798.
5. Kessner DM, Singer J, Kalk CE, Schlesinger ER. *Infant Death: An Analysis by Maternal Risk and Health Care.* Washington, DC: Institute of Medicine, National Academy of Sciences; 1973.
6. Kleinman JC, Pierre MB, Madans JH, Land GH, Schramm WF. The effects of maternal smoking on fetal and infant mortality. *Am J Epidemiol.* 1988;127:274–282.
7. Schramm W. Smoking and pregnancy outcome. *Mo Med.* 1980;77:618–626.

Infant Mortality and Related Risk Factors among Asian Americans

Hallie W. Morrow, MD, MPH, Gilberto F. Chávez, MD, MPH, Peggy P. Giannoni, MA, and Rugmini S. Shah, MD

ABSTRACT

To examine differences in perinatal health among nine Asian ethnic subgroups, a descriptive epidemiological study was conducted using linked birth/infant death certificates for 1982 to 1987. When compared with Whites, Asians had a lower proportion of young mothers, unmarried mothers, and women who received first trimester prenatal care; a higher proportion of foreign-born mothers; and a different birthweight distribution. A great deal of heterogeneity was found in risk factors and infant mortality rates among the various Asian ethnic subgroups. Paradoxically, although Asian ethnic subgroups had a higher perinatal risk profile, they had more favorable birth outcomes than did Whites. (*Am J Public Health.* 1994;84:1497–1500)

Introduction

The Asian population in the United States has grown significantly in recent years.[1,2] The large majority of Asian-American births occur in only a few states, and more than 40% of them occur in California alone. Yet we know very little about the specific perinatal health status and risk factors of the different Asian ethnic subgroups, particularly the growing Southeast Asian populations.

In the United States, Asians overall have a slightly lower infant mortality rate than do Whites (i.e., the non-Hispanic White population).[3] However, Asians do not appear to be a homogeneous group in this regard. During 1983 and 1984, reported infant mortality rates varied considerably among some of the Asian ethnic subgroups studied in California and nationwide.[3] Overall, Japanese had the lowest infant mortality rates, followed by Chinese and Filipinos.[4,5] Specific information on infant mortality rates for other Asian ethnic groups is lacking.

Our study examines the prevalence of selected risk factors, the birthweight distribution, and the infant mortality experience among nine Asian ethnic subgroups in California between 1982 and 1987.

Methods

Data for this study were obtained from the California Birth Cohort file, a computerized record of linked birth/infant death certificates. Further information on this file has been published elsewhere.[6–8] Included in the analysis were all infants delivered alive from 1982 through 1987 whose mothers were identified as belonging to any of nine Asian ethnic subgroups (Cambodian, Chinese,

The authors are with the Maternal and Child Health Branch, Primary Care and Family Health, California Department of Health Services, in Sacramento. Gilberto F. Chávez is also with the Division of Reproductive Health, National Center for Health Promotion and Chronic Disease Prevention, Centers for Disease Control and Prevention, Atlanta, Ga.

Requests for reprints should be sent to Gilberto F. Chávez, Maternal and Child Health Branch, California Department of Health Services, 714 P St, Room 300, Sacramento, CA 95814.

This paper was accepted November 18, 1993.

TABLE 1—Percentage Distribution of Maternal Risk Factors, by Race/Ethnicity: California, 1982 to 1987

		Asians									
	Whites (1 371 189)	Total (217 014)	Filipino (65 931)	Chinese (44 629)	Vietnamese (27 377)	Korean (21 195)	Laotian (20 237)	Japanese (18 396)	Cambodian (10 644)	Thai (3 098)	Other Asian (5 507)
Maternal age, y											
<15	0.1	0.1	0.1	0.0	0.1	0.0	0.6	0.0	0.1	0.1	0.1
15–17	2.7	1.4	1.5	0.2	1.1	0.2	5.6	0.8	2.0	1.2	1.4
18–19	6.0	2.8	3.2	0.5	3.5	0.8	8.0	1.7	5.1	2.1	2.5
20–29	61.8	52.1	51.1	45.0	58.9	64.5	56.0	45.5	54.8	37.2	53.5
30–34	21.3	29.8	29.4	39.4	24.7	26.9	17.1	36.5	21.1	37.2	30.1
35–39	7.2	11.5	12.7	13.6	9.7	6.7	8.2	13.7	11.3	19.1	10.9
40+	1.0	2.1	2.2	1.3	2.0	0.9	4.2	1.6	5.5	3.1	1.3
Trimester prenatal care began											
First	81.8	75.7	78.5	83.1	70.4	79.2	53.5	85.4	54.5	78.9	82.0
Second	13.6	15.9	16.2	10.1	17.3	13.2	28.3	10.1	28.3	15.0	12.0
Third	2.4	3.5	3.1	2.4	4.7	3.4	6.2	1.8	6.8	2.4	2.6
None	0.7	0.9	0.7	0.5	1.4	0.8	1.8	0.6	2.4	0.8	0.9
Unknown	1.5	4.0	1.6	3.9	6.2	3.4	10.2	2.0	8.1	2.8	2.5
Maternal birthplace											
US born	90.5	12.4	11.9	14.7	0.2	1.3	0.9	50.9	0.1	0.5	49.2
Foreign born	9.5	87.5	88.1	85.2	99.8	98.6	99.0	49.1	99.9	99.5	50.1
Marital status											
Married	83.3	92.3	88.9	96.9	91.2	98.3	92.8	92.4	85.0	93.8	90.5
Not married	16.7	7.7	11.1	3.1	8.8	1.7	7.2	7.6	15.0	6.2	9.5

Note. Excludes multiple births, non-California residents, infants under 500 g, and those with unknown maternal age and birthplace. Percentages may not add up to 100.0 because of rounding.
Source. California Department of Health Services, Birth Cohort File, 1982 to 1987.

TABLE 2—Percentage Distribution of Live Births, by Birthweight and Race/Ethnicity: California, 1982 to 1987

		Asians									
Birthweight, g	Whites (1 371 189)	Total (217 014)	Filipino (65 931)	Chinese (44 629)	Vietnamese (27 377)	Korean (21 195)	Laotian (20 237)	Japanese (18 396)	Cambodian (10 644)	Thai (3 098)	Other Asian (5 507)
<1500	0.6	0.6	0.7	0.5	0.5	0.6	0.6	0.5	0.6	0.9	0.7
1500–2499	3.6	4.6	5.5	3.8	4.6	2.9	5.0	4.3	5.6	4.7	4.2
2500–3999	81.4	88.9	87.4	89.6	91.1	87.9	89.7	88.6	90.6	88.2	87.6
4000 +	14.3	5.9	6.3	6.0	3.7	8.5	4.7	6.5	3.2	6.3	7.4
Unknown	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0

Note. Excludes multiple births, non-California residents, and infants under 500 g. Percentages may not add up to 100.0 because of rounding.
Source. California Department of Health Services, Birth Cohort File, 1982 to 1987.

Filipino, Japanese, Korean, Laotian, Thai, Vietnamese, and Other Unspecified Asians). Excluded were infants born to mothers who were not California residents at the time of delivery (0.17%), infants whose birthweights were below 500 g (0.06%), and multiple births (1.57%).

The analysis was limited to risk factors for which reliable data were available. Parity, gravidity, and gestational age were not included because the reliability of the information available for meaningful analysis was questionable. Maternal education was not collected in California during the study period. Neonatal (<28 days), postneonatal (28 days to 364 days), and infant (0 to 364 days) mortality rates per 1000 live births were calculated, as were birthweight-specific mortality rates for all groups. Relative risks and 95% confidence intervals[9] were also calculated using Whites and all Asians as reference standards when comparing outcomes. Epi-Info,[10] SAS,[11] and Syntax[12] computer programs were used for the analysis.

Results

Overall, the prevalence of teenage births was lower for Asians than for Whites (4.3% vs 8.8%) (Table 1). Chinese and Koreans had remarkably low percentages of births to teenage mothers (0.7% and 1.0%, respectively), but Laotians had a very high proportion of such births (14.2%). Asians had a higher proportion of births to mothers 35 years of age and older than did Whites (13.6% vs 8.2%). This higher prevalence of older mothers

TABLE 3—Birthweight-Specific Mortality Rates and Relative Risks among Asian Ethnic Subgroups: California, 1982 to 1987

	Birthweight, g					
	<1500	1500–2499	2500–3999	4000+	Total	RR (95% CI)
Neonatal						
Whites	268.0	21.0	1.5	1.0	4.0	1.00 (Referent)
Asians	251.5	18.2	1.3	1.5	3.9	0.91 (0.85, 0.98)
Filipino	236.1	11.0	1.7	2.6	4.2	0.87 (0.77, 0.98)
Chinese	259.1	22.4	1.0	...	3.2	0.87 (0.74, 1.02)
Vietnamese	310.1	26.0	1.2	...	4.1	1.06 (0.89, 1.27)
Korean	273.4	21.4	1.6	...	4.1	1.01 (0.82, 1.25)
Laotian	235.8	26.7	1.7	...	4.5	1.06 (0.86, 1.29)
Japanese	182.5	15.2	0.8	...	2.7	0.62 (0.47, 0.84)
Cambodian	220.3	15.1	1.0	...	3.1	0.74 (0.53, 1.03)
Thai	481.5	...	...	...	6.5	1.37 (0.94, 2.01)
Other Asian	236.8	34.5	1.0	...	4.2	0.97 (0.65, 1.44)
Postneonatal						
Whites	49.8	12.8	2.7	1.7	3.3	1.00 (Referent)
Asians	47.5	9.6	2.2	1.2	2.8	0.80 (0.73, 0.87)
Filipino	43.1	8.0	2.0	...	2.6	0.72 (0.62, 0.84)
Chinese	59.1	13.6	2.3	1.9	3.0	0.92 (0.78, 1.09)
Vietnamese	38.8	8.7	2.0	...	2.5	0.72 (0.57, 0.92)
Korean	46.9	...	2.0	2.8	2.4	0.75 (0.57, 0.98)
Laotian	89.4	9.9	3.0	...	3.7	1.09 (0.87, 1.37)
Japanese	...	8.8	2.0	...	2.4	0.68 (0.50, 0.93)
Cambodian	...	10.1	1.8	...	2.4	0.67 (0.45, 0.99)
Thai	...	...	2.6	...	2.9	0.82 (0.43, 1.57)
Other Asian	...	21.6	2.1	...	3.3	0.96 (0.61, 1.52)
Infant						
Whites	317.8	33.8	4.2	2.7	7.3	1.00 (Referent)
Asians	299.1	27.9	3.5	2.7	6.6	0.86 (0.82, 0.91)
Filipino	279.3	19.0	3.7	3.6	6.8	0.81 (0.74, 0.89)
Chinese	318.2	36.0	3.3	3.0	6.3	0.90 (0.80, 1.00)
Vietnamese	348.8	34.7	3.2	...	6.6	0.90 (0.78, 1.04)
Korean	320.3	26.4	3.5	3.3	6.5	0.89 (0.75, 1.05)
Laotian	325.2	36.6	4.7	...	8.3	1.06 (0.91, 1.23)
Japanese	204.3	24.0	2.8	...	5.1	0.65 (0.53, 0.80)
Cambodian	237.3	25.2	2.8	...	5.5	0.71 (0.55, 0.91)
Thai	481.5	...	3.7	...	9.4	1.13 (0.81, 1.58)
Other Asian	315.8	56.0	3.1	...	7.4	0.96 (0.72, 1.29)

Note. Excludes multiple births, non-California residents, infants with unknown birthweights, and infants under 500 g. Mortality rates are per 1000 live births; relative risks (RRs) and confidence intervals (CIs) are birthweight adjusted. Rates not presented for cells with fewer than five events.
Source. California Department of Health Services, Birth Cohort File, 1982 to 1987.

was true for all the Asian subgroups except Koreans (7.6%). Additionally, more Asian mothers than White mothers were giving birth after the age of 39. Laotians and Cambodians, the groups with the highest percentage of teenage births, also had the highest percentage of births among women 40 years of age and older.

Compared with Whites, Asians had a lower proportion of women who received prenatal care in the first trimester of pregnancy and a higher proportion of women who received no prenatal care (Table 1). Only about 50% of Laotian and Cambodian women received prenatal care in the first trimester. In addition, these groups had the highest percentages of women who received no prenatal care and whose prenatal care history was unknown.

Asian mothers overall were nine times more likely than Whites to have been born outside the United States (Table 1). In several of the Asian subgroups—Vietnamese, Korean, Laotian, Cambodian, and Thai—essentially all the mothers were women who had immigrated to this country.

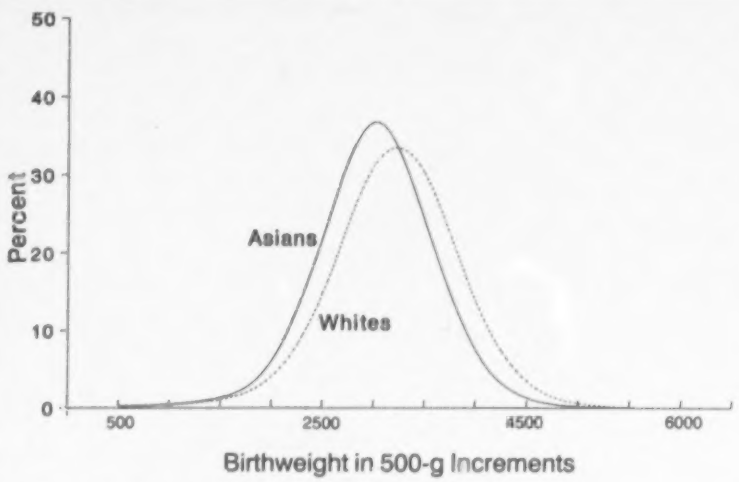
As a group, Asian mothers were less likely to be unmarried than White mothers (7.7% vs 16.7%) (Table 1). Each of the specific Asian subgroups also had a lower prevalence of unmarried mothers than Whites did. Koreans and Chinese had the lowest prevalence of unmarried mothers (1.7% and 3.1%, respectively); Filipinos and Cambodians had the highest prevalence (11.1% and 15.0%, respectively).

There were more low-birthweight infants (<2500 g) born to Asians than to Whites (5.2% vs 4.2%) (Table 2). This was true for all the Asian subgroups except Chinese, for whom there was no difference, and Koreans, who had fewer low-birthweight infants than Whites did. Examining the differences between all Asians combined and Whites by 500-g increments, we observed that the birthweight distribution for Asians was different than that for Whites and appeared to be shifted toward lower birthweight infants (Figure 1).

Asians as a group and the Filipino, Japanese, and Cambodian subgroups had significantly lower infant mortality risks than Whites (Table 3). The Thai and Laotian subgroups had slightly, but not statistically significant, elevated risks for infant mortality. Neonatal mortality was significantly lower for Asians as a group and for the Filipino and Japanese subgroups. The risks for postneonatal mortality were significantly lower for Asians as a group and for the Filipino, Vietnamese, Korean, Japanese, and Cambodian subgroups. After adjusting for birthweight, we found that Asians had lower overall neonatal, postneonatal, and infant mortality rates than did Whites.

Discussion

We showed a high degree of variation among the various Asian ethnic groups for maternal risk factors. Additionally, we observed that ethnic-specific differences in birthweight distributions, which have been reported for Chinese and Southeast Asian infants,[13–15] exist overall among several Asian ethnic subgroups when compared with Whites. We believe that our most remarkable finding was that even with great variability in the frequency of risk factors among the various Asian subgroups, the variability in birth outcomes is relatively small. In addition, despite their higher frequency of risk factors, Asians overall had significantly better birth outcomes than did Whites. Many factors could contribute to this finding, including differences in education, acculturation, nutrition, access to and content of prenatal care, life-style, family structure, social support, and underreporting. This apparent epidemiological paradox deserves further investigation since much may be learned from those ethnic groups that, despite serious disad-



Note. Excludes multiple births, non-California residents, and infants under 500 g.

FIGURE 1—Percentage distribution of live births by birthweight and race/ethnicity: California, 1982 to 1987.

vantages, are able to have healthier babies.

The high proportion of foreign-born Asian mothers in our study emphasizes the need for developing culturally appropriate multilingual prenatal care and health care services for this growing population. Larger sample sizes are needed to examine the differences in infant mortality risks between US- and foreign-born mothers with respect to acculturation and socioeconomic status.

Our study emphasizes the need for continued analysis of ethnic-specific perinatal health data to assess community needs adequately and to target health education efforts and intervention strategies to the groups at highest risk for suboptimal pregnancy outcomes and to those with the highest need for services. □

Acknowledgments

We thank Donna Steen for providing assistance in programming, Celine Donaldson for preparing the tables and graph, Mich Tashiro, and Penelope Stephenson. We are also indebted to Susan Chu, PhD, Jose Becerra, MD, and Ray Yip, MD, of the Centers for Disease Control and Prevention for reviewing the manuscript.

References

1. National Center for Health Statistics. Advance report of final natality statistics, 1982. *Month Vital Stat Rep.* 1984;33(6 suppl).
2. National Center for Health Statistics. Advance report of final natality statistics, 1988. *Month Vital Stat Rep.* 1990;39(4 suppl).
3. Kleinman JC. Infant mortality among racial/ethnic minority groups, 1983–1984. *MMWR.* 1990;39(SS-3):31–39.
4. Tashiro M, Oreglia A. *Infant Death Rates by Race/Ethnicity, California Birth Cohorts, 1978–1987.* Sacramento, Calif: State of California, Department of Health Services; August 1990. Report Register 90–05002.
5. Oreglia A. *California's Infant Death Rate—1989.* Sacramento, Calif: State of California, Department of Health Services; February 1991. Report Register 91–02000.
6. Oreglia A, Tashiro M. *What Are California's Birth Cohort Files? Questions and Answers.* Sacramento, Calif: State of California, Department of Health Services; November 1986. Report Register 86–10144.
7. Tashiro M. *A Description of the California Birth Cohort Perinatal File.* Sacramento, Calif: State of California, Department of Health Services; February 1984. Report Register 83–11078.
8. Williams RL, Rust FP, Chen PM, Waller S. *1982–1986 Maternal and Child Health Data Base: Descriptive Narrative.* Santa Barbara, Calif: Regents of the University of California; 1990.
9. Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. *Biometrics.* 1985;41:55–68.
10. Dean AD, Dean JA, Burton JH, Dicker RC. *Epi-Info, Version 5: A Word Processing, Database, and Statistics Program for Epidemiology on Microcomputers.* Atlanta, Ga: Centers for Disease Control; 1990.
11. *SAS User's Guide: Basics, Version 5 Edition.* Cary, NC: SAS Institute Inc; 1985.
12. *Syntax II (Tables and Editor) Documentation–Version V.O.* Berkeley, Calif: University of California; 1977.
13. Yip R, Li Z, Chong WH. Race and birth weight: the Chinese example. *Pediatrics.* 1991;87:688–693.
14. Helsel D, Petitti DB, Kunstadter P. Pregnancy among the Hmong: birthweight, age, and parity. *Am J Public Health.* 1992;82: 1361–1364.
15. Gann P, Nghiem L, Warner S. Pregnancy characteristics and outcomes of Cambodian refugees. *Am J Public Health.* 1989;79: 1251–1257.

The Perinatal and Infant Health Status of Native Hawaiians

Edith C. Kieffer, PhD, MPH, Joanne M. Mor, MS, and Greg R. Alexander, MPH, ScD

ABSTRACT

Hawaii vital record data for 1979 through 1990 were analyzed to examine potentially differing relationships between maternal and infant risks and outcomes in native Hawaiian and White infants. Despite high rates of inadequate prenatal care and teenage and unmarried childbearing, the Hawaiian low-birthweight rate was below the US average. Hawaiian infants experienced an elevated risk of mortality, particularly among those of normal birthweight during the postneonatal period. Public health initiatives to reduce infant mortality must go beyond preventing teenage pregnancy and low birthweight to address Hawaiian infants' unique pattern of risk factors and the social and economic environment in which such risks abound. (*Am J Public Health.* 1994;84: 1501–1504)

Introduction

The indigenous people of Hawaii experience higher rates of morbidity and mortality from chronic diseases than do other ethnic groups in Hawaii and the United States.[1–5] Risks for diabetes and hypertension are particularly high among Native Hawaiian women and increase with age.[3,4] Less is known about the health status of Hawaiian infants. Recent reports have noted high rates of teen pregnancy and unmarried childbearing but relatively low rates of low birthweight among Hawaiian infants.[6,7] Similar patterns have been noted for American Indian and Mexican-American populations.[7,8] Such findings have stimulated discussion of protective factors that may offset the presumed higher risk of low birthweight.[9–11]

This study examined birthweight and gestational age, as well as rates and causes of mortality during the first year of life, for Hawaiian infants. Variations in maternal characteristics and health status indicators between Hawaiian and White infants born in Hawaii were investigated to identify the potentially differing role of maternal sociodemographic risk factors in these populations.

Data and Methods

Data were derived from Hawaii's linked live birth–infant death vital record files for 1979 through 1990. Data for the 12-year period were combined to minimize unstable fluctuations associated with small numbers of events and unusual, short-term trends. There were no appreciable temporal trends in birth outcomes over the period. All single live births to Hawaii resident mothers who identified themselves on the birth certificate as of Hawaiian, part-Hawaiian, or Caucasian ancestry were included in the study population. In this analysis, the category Hawaiian includes mothers with any Hawaiian ancestry, following common practice in the state.[12] An estimated 2% to 5% of these women are of unmixed Hawaiian ancestry.[5] A total of 117 005 births, including 66 840 White and 50 165 Hawaiian infants, were analyzed.

After imputing for missing values,[13] gestational age data, based on date of last normal menses, were available for 96.8% of the cases. Because gestational ages are typically less reliably reported, caution should be used in their interpretation. Less than 5% of the data for the other variables in the study were missing.

The significance of the bivariate relationships between maternal characteristics and infant outcomes was tested with chi-square and *t*-test statistics. Multiple logistic regression analyses were used to estimate associations between the independent effects of maternal characteristics and infant outcomes.[14]

Results

Maternal characteristics are defined and summarized in Table 1. Forty percent of Hawaiian mothers were unmarried, in comparison with 13% of White mothers. Hawaiian women had their first births at earlier ages than White mothers and had higher parities at each age level. Hawaiian women were four times more likely to have their first child, and three times more likely to have one or more previous children, before they were 18 years old. White women were more likely to have their first child after 29 years of age. Hawaiian women had lower levels of educational attainment, and only 56% received adequate levels of prenatal care. More than half of Hawaiian mothers resided in rural areas, in comparison with

Edith C. Kieffer is with the School of Public Health, University of Michigan, Ann Arbor. Joanne M. Mor is with the School of Public Health, University of Hawaii at Manoa. Greg R. Alexander is with the School of Public Health, University of Minnesota, Minneapolis.

Requests for reprints should be sent to Edith C. Kieffer, PhD, MPH, University of Michigan, School of Public Health, 1420 Washington Heights, Ann Arbor, MI 48109-2029.

This paper was accepted January 26, 1994.

TABLE 1—Maternal Sociodemographic Characteristics of Hawaiian and White Women: Hawaii Resident Single Live Births, 1979 through 1990

Maternal Characteristic	Maternal Ethnicity: Hawaiian,[a] % (No.)	Maternal Ethnicity: White % (No.)	Chi-Square P
Unmarried	40.0 (20 080)	13.3 (8 841)	<.01
Age <18 y, para 0	5.6 (2 820)	1.4 (955)	
Age <18 y, para ≥1	0.7 (356)	0.2 (97)	
Age 18–29 y, para 0	29.6 (14 823)	36.1 (24 103)	
Age 18–29 y, para 1–4	45.5 (22 813)	33.3 (22 233)	<.01
Age 18–29 y, para ≥5	1.1 (530)	0.1 (83)	
Age ≥30 y, para 0	2.4 (1 191)	9.0 (6 038)	
Age ≥30 y, para 1–4	13.3 (6 647)	19.2 (12 822)	
Age ≥30 y, para ≥5	1.9 (972)	0.7 (497)	
Low educational attainment[b]	15.3 (7 639)	7.6 (5 075)	<.01
High educational attainment[c]	22.6 (11 295)	47.0 (31 332)	<.01
Received adequate prenatal care[d]	55.8 (26 291)	71.6 (46 049)	<.01
Not born in Hawaii	6.1 (3 039)	88.0 (58 781)	<.01
Rural residence[e]	54.0 (27 016)	32.1 (21 398)	<.01

Note. Analyses included 50 165 Hawaiian and 66 838 White infants. Percentages do not include missing values.
[a]Includes part-Hawaiians.
[b]2 or more years below expected grade level for age for adolescents (less than 18 years) or less than 12 years for adults.
[c]2 or more years above expected grade level for age for adolescents (less than 18 years) or more than 12 years for adults.
[d]Based on Kessner et al.[18]
[e]Includes the islands of Hawaii, Maui, Molokai, Lanai, Kauai, and Niihau and the Waianae Coast, North Shore, North Windward, and Waimanalo areas of Oahu.

TABLE 2—Birth Outcomes of Infants Born to Hawaiian and White Women: Hawaii Resident Single Live Births, 1979 through 1990

Birth Outcome	Maternal Ethnicity: Hawaiian	Maternal Ethnicity: White	P
Low birthweight (<2500 g), % (No.)	6.6 (3 323)	4.8 (3 193)	<.01
Very low birthweight (<1500 g), % (No.)	1.0 (501)	0.7 (465)	<.01
High birthweight (≥4250 g), % (No.)	3.4 (1 717)	5.2 (3 478)	<.01
Very high birthweight (≥4500 g), % (No.)	1.3 (653)	1.9 (1 927)	<.01
Birthweight, g, mean ± SD	3285 ± 563.8	3410 ± 556.3	<.01
Preterm (<37 wk), % (No.)	10.3 (4 954)	6.6 (4 263)	<.01
Very preterm (<33 wk), % (No.)	2.3 (1 103)	1.2 (783)	<.01
Gestational age, wk, mean ± SD	39.3 ± 2.80	39.7 ± 2.45	<.01
Small for gestational age,[a] % (No.)	6.3 (2 959)	4.6 (2 943)	<.01
Large for gestational age,[a] % (No.)	17.7 (8 371)	20.1 (12 837)	<.01
Neonatal mortality rate	6.2 (310)	4.3 (290)	<.01
Postneonatal mortality rate	3.8 (190)	1.9 (130)	<.01
IMR	10.0 (500)	6.3 (420)	<.01
Infectious disease–specific IMR	0.6 (30)	0.3 (20)	.02
Injury-specific IMR	0.3 (17)	0.1 (8)	<.01
Congenital anomaly–specific IMR	2.4 (122)	2.1 (139)	>.10
Perinatal condition–specific IMR (No.)	4.3 (214)	2.5 (170)	<.01
SIDS-specific IMR	1.6 (80)	0.8 (53)	<.01
Other cause IMR	0.6 (32)	0.4 (28)	>.10

Note. Rates are per 1000 live births. IMR = infant mortality rate; SIDS = sudden infant death syndrome.
[a]Based on Brenner.[19]

a third of White mothers. Almost 90% of White mothers but only 6% of Hawaiian mothers were born outside of Hawaii. Because of extreme collinearity, this variable was excluded from subsequent analyses.

Table 2 summarizes measures of birthweight, gestational age, and infant mortality. The low-birthweight percentage for Hawaiian infants was 6.6, with very-low-birthweight infants accounting for 1% of births. Hawaiian percentages were below the 1990 US average but were significantly higher than rates for White infants.[7] Hawaiian infants were less likely to be born at high birthweights, and their mean birthweight was 125 g lower than that of White infants. Hawaiian infants were significantly more likely to be preterm and had almost double the risk of very preterm birth. Hawaiian infants were more likely to be small for gestational age and had slightly but significantly shorter mean gestational age.

Hawaiian infants experienced significantly higher infant mortality, including a postneonatal mortality rate that was double that of White infants. Higher infant mortality rates for Hawaiians were observed for every major cause of death. Perinatal conditions accounted for the greatest proportion of deaths to Hawaiian infants, followed by congenital anomalies and sudden infant death syndrome. Mortality rates from infectious diseases and injuries were two and three times the rates for White infants, respectively.

When birthweight and gestational age-specific infant mortality rates were examined, the mortality differential between ethnic groups was significant only for infants of normal birthweight and gestational age (Table 3). Differences between Hawaiian and White infants in rates of birthweight and gestational age-specific neonatal mortality were not significant. However, Hawaiian infants of moderately low birthweight and normal birthweight were almost twice as likely to die in the postneonatal period as White infants in the same birthweight categories. Postneonatal mortality rates were significantly higher for Hawaiian infants of normal gestational age.

Hawaiian infants remained at increased risk of low birthweight, preterm delivery, and neonatal and postneonatal mortality after logistic regression analysis accounted for the effects of maternal sociodemographic and prenatal care use characteristics (Table 4).

Discussion

This study contributes additional evidence of the discrepancy in health status between indigenous people of the United States and more recent residents.[8,15] The most common measure of

TABLE 3—Birthweight and Gestational Age–Specific Infant Mortality Rates,[a] by Maternal Ethnicity: Hawaii Resident Single Live Births, 1979 through 1990

	Neonatal Mortality Rate (No.)			Postneonatal Mortality Rate (No.)			Infant Mortality Rate (No.)		
	Hawaiian	White	*P*	Hawaiian	White	*P*	Hawaiian	White	*P*
Birthweight, g									
<1500	365.3 (183)	316.1 (147)	.11	41.9 (21)	36.6 (17)	.67	407.2 (204)	352.7 (164)	.08
1500–2499	15.9 (45)	17.6 (48)	.63	11.3 (32)	6.2 (17)	.04	27.3 (77)	23.8 (65)	.42
2500–4249	1.4 (63)	1.3 (80)	.77	3.0 (133)	1.6 (94)	<.01	4.3 (196)	2.9 (174)	<.01
≥4250	1.7 (3)	1.2 (4)	.69[b]	1.2 (2)	0.6 (2)	.60[b]	2.9 (5)	1.7 (6)	.52[b]
Gestational age, wk									
<33	135.1 (149)	163.5 (128)	.09	19.9 (22)	16.6 (13)	.60	155.0 (171)	180.1 (141)	.15
33–36	7.3 (28)	10.6 (37)	.13	4.7 (18)	2.6 (9)	.14	11.9 (46)	13.2 (46)	.63
37–41	1.9 (68)	1.4 (71)	.09	3.1 (113)	1.7 (88)	<.01	5.0 (181)	3.1 (159)	<.01
≥42	1.7 (12)	1.8 (19)	.84	3.6 (25)	1.6 (16)	.01	5.3 (37)	3.4 (35)	.06

[a]Rates are per 1000 live births.
[b]Based on Fisher's exact test because of expected cell values less than 5.

TABLE 4—Logistic Regression Analysis of Factors Associated with Birth Outcomes: Hawaii Resident Single Live Births, 1979 through 1990

	Odds Ratio (95% Confidence Interval)			
	Low Birthweight	Preterm	Neonatal Mortality	Postneonatal Mortality
Age <18 y, para 0	1.36 (1.19, 1.56)	1.38 (1.24, 1.53)	1.56 (1.01, 2.41)	0.39 (0.19, 0.81)
Age <18 y, para ≥1	1.70 (1.23, 2.35)	1.75 (1.34, 2.28)	4.42 (2.14, 9.13)	0.43 (0.06, 3.09)
Age 18–29 y, para 0	1.38 (1.30, 1.48)	1.07 (1.02, 1.13)	1.31 (1.05, 1.63)	0.73 (0.55, 0.95)
Age 18–29 y, para ≥5	0.99 (0.69, 1.43)	1.34 (1.04, 1.74)	2.21 (0.90, 5.44)	1.66 (0.61, 4.55)
Age ≥30 y, para 0	1.98 (1.77, 2.22)	1.35 (1.22, 1.49)	1.35 (0.90, 2.02)	0.85 (0.48, 1.51)
Age ≥30 y, para 1–4	1.16 (1.07, 1.27)	1.12 (1.05, 1.20)	0.91 (0.67, 1.24)	0.58 (0.38, 0.87)
Age ≥30 y, para ≥5	1.13 (0.89, 1.44)	1.23 (1.02, 1.48)	0.92 (0.38, 2.25)	1.31 (0.57, 2.97)
Unmarried	1.32 (1.23, 1.40)	1.32 (1.25, 1.39)	1.22 (0.98, 1.52)	2.00 (1.54, 2.60)
High educational attainment	0.79 (0.74, 0.85)	0.87 (0.82, 0.92)	0.92 (0.74, 1.14)	0.89 (0.66, 1.19)
Low educational attainment	1.43 (1.33, 1.55)	1.20 (1.12, 1.28)	0.80 (0.59, 1.10)	1.31 (0.96, 1.80)
Received adequate prenatal care	0.73 (0.69, 0.77)	0.71 (0.67, 0.74)	0.79 (0.65, 0.95)	0.81 (0.63, 1.02)
Rural residence	0.95 (0.89, 1.00)	0.95 (0.90, 0.99)	1.01 (0.83, 1.21)	0.93 (0.73, 1.18)
Hawaiian	1.18 (1.11, 1.26)	1.38 (1.31, 1.45)	1.23 (1.01, 1.51)	1.42 (1.10, 1.85)

Note. The reference group was as follows: age 18–29 years, para 1–4, married, average educational attainment, received subadequate prenatal care, urban residence, White (see Tables 1 and 2 for definitions).

infant health status, low birthweight, is not excessively high in this population. It is difficult to assess the public health significance of the modest increase in risk among Hawaiian infants for low birthweight, particularly given the large number of cases in this analysis. However, neonatal and postneonatal mortality rates are higher than would be expected given the percentage of low birthweight. Indeed, the ethnic gap in mortality between White and Hawaiian, American Indian, and Black infants is widest for infants born at normal birthweights and gestational ages.[8,15]

Common maternal risk factors, including age, parity, education, and prenatal care use, did not account for the excess risk of Hawaiian infant mortality. While many Hawaiian mothers reside in geographically isolated rural areas that have poor access to health care services, rural residence was not a significant factor in explaining ethnic differences in infant health status.

Maternal height and weight, tobacco and alcohol use, and other measures of maternal poverty and chronic disease were not available for the study population. Data from 1988 through 1990 birth certificates indicate that 20% of Hawaiian and 13% of White mothers reported smoking during pregnancy, a discrepancy that, if accurate, could account for some of the variation in birthweight between Hawaiian and White infants. Previous studies have noted that the mean height of Hawaiian mothers approximates that of White mothers in Hawaii, while mean weight and weight gain in pregnancy are higher.[16] Given the high prevalence of hypertension, diabetes, and asthma among Hawaiian women of childbearing age, the possible role of chronic diseases in the pregnancy outcomes of Hawaiian women deserves attention.

Higher Hawaiian postneonatal mortality rates are congruent with classic public health theories relating the risk of postneonatal death to poverty and environmental hazards. The increased risk of mortality from such preventable causes as infectious disease and injury among Hawaiian infants has been documented

previously.[17] Yet, intensive public health initiatives to reduce infant mortality in Hawaii and elsewhere have focused on the reduction of teenage pregnancy and low birthweight. High parity among older Hawaiian women and higher rates of mortality for Hawaiian infants of normal birthweight and gestational age suggest that improvement of maternal and infant health status for the indigenous people of Hawaii will necessitate a better understanding of their unique pattern of risk factors and interventions that address the social and economic environment in which such risks abound. □

Acknowledgments

We thank the Office of Health Status Monitoring of the Hawaii State Department of Health for assistance in obtaining the data for this study and the Maternal and Child Health Program, School of Public Health, University of Hawaii for providing support for this research.

References

1. Wegner E. Preface. *Soc Process Hawaii.* 1989;32(special issue):xi–xvi.
2. Johnson D. An overview of ethnicity and health in Hawaii. *Soc Process Hawaii.* 1989;32(special issue):67–86.
3. Johnson D. Diabetes: epidemiology and disability. *Soc Process Hawaii.* 1989;32(special issue):104–112.
4. Wegner E. Hypertension and heart disease. *Soc Process Hawaii.* 1989;32(special issue):113–133.
5. Miike L. *Current Health Status and Population Projections of Native Hawaiians Living in Hawaii.* Washington, DC: Office of Technology Assessment; 1987.
6. Park C, Horiuchi B. Ethnicity, birth weight and maternal age in infant mortality: Hawaiian experience. *Am J Hum Biol.* 1993;5:101–109.
7. Centers for Disease Control and Prevention. Advance report of final natality statistics, 1990. *Month Vital Stat Rep.* February 25, 1993; 41(9) suppl:1–52.
8. Kleinman J. Infant mortality among racial/ethnic minority groups, 1983–1984. *MMWR.* 1990;39(suppl 3):31–39.
9. Scribner R, Dwyer J. Acculturation and low birth weight among Latinos in the Hispanic HANES. *Am J Public Health.* 1989;79:1263–1267.
10. Kieffer E, Alexander G, Lewis N, Mor J. Geographic patterns of low birth weight in Hawaii. *Soc Sci Med.* 1993;36:557–564.
11. James S. Racial and ethnic differences in infant mortality and low birth weight: a psychosocial critique. *Ann Epidemiol.* 1993; 3:130–136.
12. Aiu P, Miike L. *Native Hawaiian Health Data Book 1992.* Honolulu, Hawaii: Papa Ola Lokahi; 1992.
13. Alexander GR, Tompkins ME, Cornely DA. Gestational age reporting and preterm delivery. *Public Health Rep.* 1990;105: 267–275.
14. *SAS/STAT Users Guide, Version 6, 4th Edition, Vol. II.* Cary, NC: SAS Institute Inc; 1989.
15. Honigfeld L, Kaplan D. Native American postneonatal mortality. *Pediatrics.* 1987;80: 575–578.
16. Werner E, Bierman J, French F. *The Children of Kauai.* Honolulu, Hawaii: University of Hawaii Press; 1971.
17. Baruffi G, Alexander G, Novotny R. Causes of infant mortality in a multiethnic population. *Asia-Pacific J Public Health.* 1990;4:145–150.
18. Kessner D, Singer J, Kalk C, Schlesinger E. Infant death: An analysis of maternal risk and health care. In: *Contrasts in Health Status.* Washington, DC: Institute of Medicine, National Academy of Sciences; 1973; 1.
19. Brenner W, Edelman D, Hendricks C. A standard of fetal growth for the United States of America. *Am J Obstet.* 1976;126: 555–564.

Call for Student Papers on Preventive Medicine

The American College of Preventive Medicine, the Association of Teachers of Preventive Medicine Foundation, the Association of American Medical Colleges, the Association of Preventive Medicine Residents, the journal *Preventive Medicine,* and the Ulrich and Ruth Frank Foundation for International Health announce their annual competition for the best paper in preventive medicine by a medical student. The prize for the best paper is $1000 and publication in *Preventive Medicine.* The awardee will also be honored at Prevention '95, the annual meeting of the American College of Preventive Medicine and the Association of Teachers of Preventive Medicine, to be held in New Orleans, La, in spring 1995.

All prevention-oriented topics will be considered; papers addressing the implementation of clinical preventive services are encouraged.

All submissions must be received by *November 8, 1994.* For more information, send a self-addressed, stamped envelope to Mirza I. Rahman, MD, MPH, Department of Preventive Medicine, HSC Level 3 Room 086, Stony Brook School of Medicine, Stony Brook, NY 11794-8036.

Pregnancy Risk Factors and Birth Outcomes in Washington State: A Comparison of Ethiopian-Born and US-Born Women

Haimanot Wasse, MPH, Victoria L. Holt, PhD, and Janet R. Daling, PhD

ABSTRACT

This study compared birth outcomes of three cohorts of women with singleton live births in Washington State between 1980 and 1991: all Black Ethiopian-born women (n = 264) and samples of US-born Black (n = 526) and White (n = 546) women. Ethiopians were older, more likely to be married, and less likely to smoke than were US-born women, and they were less likely to have anemia than US-born Blacks. The adjusted low-birthweight risk among Ethiopians was similar to that of US-born Blacks (relative risk [RR] = 0.9) and Whites (RR = 1.1). Ethiopians were more likely to have high-birthweight infants than were US-born Blacks (RR = 4.0). Cultural and behavioral etiologies may explain these favorable birth outcomes. (*Am J Public Health.* 1994;84: 1505–1507)

Introduction

Blacks born outside of the United States have generally been found to have more favorable pregnancy outcomes in America than their US-born counterparts, with lower rates of prematurity and higher birthweights.[1–5] The research producing these results, primarily studies of Caribbean-born immigrants of high socioeconomic status and educational levels, has not included foreign-born Black women arriving under refugee circumstances. It is unclear whether these more favorable outcomes occur as well in Black refugee populations, which may be more similar in socioeconomic status to US-born Blacks.

To address this question, we compared birth outcomes of Ethiopian-born women delivering infants in Washington State with those of US-born Black and White women. Famine, displacement, and civil war in Ethiopia in the mid-1980s led to greatly increased emigration, and most Ethiopian women in the United States are likely to have arrived recently under refugee circumstances.[6,7]

Methods

Birth certificates for all Washington State singleton births between 1980 and 1991 to Black women born in Ethiopia (n = 264) were obtained and matched in a 1:2 ratio by delivery year to births to randomly selected US-born White (n = 546) and Black (n = 532) Washington women.

Birth certificate data were used to examine demographic differences among these cohorts, and the association between maternal birthplace and infant birthweight was calculated with the Mantel-Haenszel pooled estimator of relative risk (RR) and Greenland/Robins 95% confidence intervals (CIs). Several potential confounders of this association (age, smoking, marital status, parity, prenatal care, and diabetes) were considered, and those variables that altered the relative risk by 10% or greater were included as confounders. US census data were used to calculate family income, defined as the median income of the census tract of residence, for urban residents only.

Results

Births in Washington State to Ethiopian-born women increased from 16 in 1980 through 1982 to 136 in 1989 through 1991. Ethiopian women were older, more likely to be married, less likely to smoke, and less likely to be primiparous than US-born Black and White women (Table 1). The family income of both urban Ethiopians and US-born urban Black women was substantially lower than that of urban White women. Ethiopians and US-born Blacks were less likely than Whites to initiate prenatal care early (Table 2). Gestational diabetes and preeclampsia were more common among Ethiopians than among both US-born groups, while chronic hypertension was less frequent. Anemia during pregnancy occurred among Ethiopian women half as often as among US-born Black women.

Ethiopian-born women were less likely than US-born Blacks to have low-birthweight infants (5.3% vs 7.6%). After adjustment for smoking, marital status, and prenatal care, the relative risk of low birthweight among Ethiopians was similar to that of US-born Blacks (RR = 0.9, 95% CI = 0.5, 1.7) and Whites (RR = 1.1, 95% CI = 0.6, 2.1). Twenty percent of infants born to Ethiopian women weighed 4000 g or more, in comparison with 4% of

Haimanot Wasse is with the Department of Health Services and Victoria L. Holt is with the Departments of Health Services and Epidemiology, School of Public Health and Community Medicine, University of Washington, Seattle. Janet R. Daling is with the Fred Hutchinson Cancer Research Center, Seattle.

Requests for reprints should be sent to Victoria L. Holt, PhD, Department of Health Services, SC-37, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195.

This paper was accepted March 22, 1994.

TABLE 1—Demographic and Behavioral Characteristics of Cohorts Giving Birth in Washington State, 1980 through 1991

	Ethiopian (n = 264)	Black (n = 532)	White (n = 546)
Maternal age, no. (%)			
<20 y	6 (2.3)	112 (21.1)	48 (8.8)
20–29 y	158 (60.3)	335 (63.0)	337 (61.7)
30+ y	98 (37.4)	85 (16.0)	161 (29.4)
Unknown	0 (0.0)	2 (0.8)	0 (0.0)
Marital status, no. (%)			
Married	240 (90.9)	261 (49.1)	438 (80.7)
Unmarried	24 (9.1)	271 (50.9)	105 (19.3)
Unknown	0 (0.0)	0 (0.0)	3 (0.6)
Smoking during pregnancy, no. (%)			
No	174 (97.8)	296 (78.1)	356 (77.6)
Yes	4 (2.2)	83 (21.9)	103 (22.4)
Unknown	86 (32.6)	153 (28.8)	87 (15.9)
Family income,[a] dollars	25 971	25 483	35 640
Urban residence,[b] no. (%)	240 (91.0)	452 (85.0)	256 (47.0)
Parity, no. (%)			
0	94 (36.0)	230 (43.6)	216 (40.1)
1	86 (33.0)	147 (27.9)	193 (35.9)
2+	81 (31.1)	150 (28.5)	129 (23.9)
Unknown	3 (1.1)	5 (0.9)	8 (1.5)

[a]Among urban residents, measured as median income of census tract.
[b]Defined as giving birth in King, Pierce, or Snohomish counties.

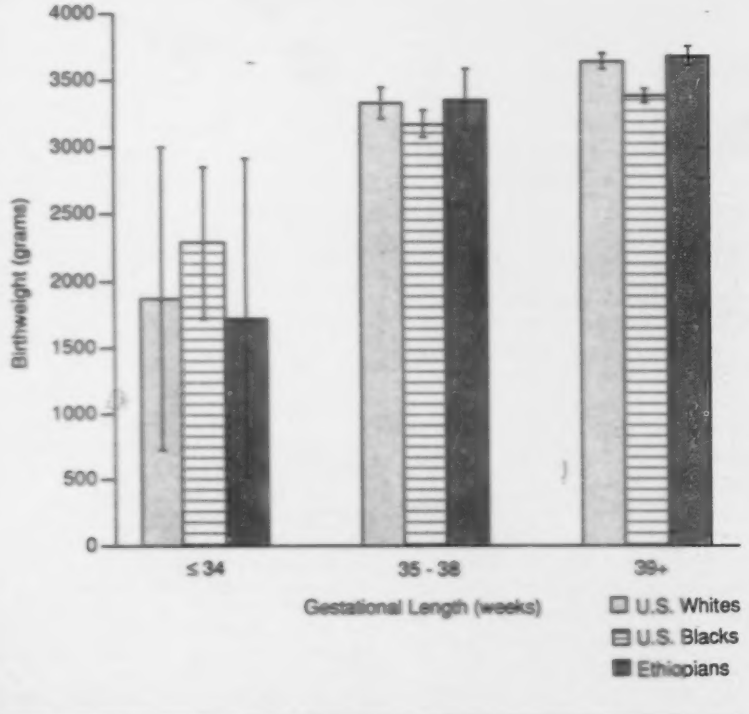


FIGURE 1—Mean birthweights (and 95% confidence intervals), by gestational age and race/ethnicity among nonsmokers.

US-born Black women's infants. After adjustment for smoking, maternal age, and marital status, Ethiopians maintained a significant elevation of macrosomia risk (RR = 4.0, 95% CI = 2.3, 6.8) relative to Black women; however, they evidenced no excess risk (RR = 1.1, 95% CI = 0.8, 1.4) when compared with Whites. Among nonsmokers, Ethiopians had infants of significantly higher mean birthweight than Blacks above 38 weeks gestation ($P < .0001$) (Figure 1).

Discussion

The interpretation of our findings is potentially limited by the absence of detailed socioeconomic information, making it possible that the favorable birth outcomes among Ethiopians in comparison with US-born Blacks were due to socioeconomic advantages. However, reports describing Ethiopian refugees indicate that, unlike the more urban, educated Ethiopians entering the United States in the early 1980s, many of those entering during the mid-1980s and later were illiterate and poor, had little formal education or skills, and were from rural areas.[7,8] The dramatic increase in Washington State deliveries among Ethiopians during the mid- and late 1980s, when a large population of Ethiopian refugees was displaced and left Ethiopia, suggests that many, if not most, of the Ethiopian women in our study were refugees.[9,10] In addition, we found that family income, measured by median income of census tract of residence, was similar among Ethiopians and US-born Blacks, making it unlikely that Ethiopians were an advantaged group.

That Ethiopian women had no greater risk of delivering low-birthweight infants than US-born Whites is surprising in light of their lack of prenatal care and the difficulties they may face upon entering the country. Even more surprising, however, is that 20% of Ethiopian women delivered infants weighing more than 4000 g. The high proportion of large infants born to Ethiopians in comparison with US-born Blacks was not associated with differences in parity, gestational age, or gestational diabetes. Furthermore, among term deliveries, the mean birthweight of Ethiopians was more similar to that of US-born Whites than to that of US-born Blacks. The higher apparent mean birthweight among Blacks at less than 35 weeks gestation shown in Figure 1

TABLE 2—Pregnancy and Neonatal Characteristics of Cohorts Giving Birth in Washington State, 1980 through 1991

	Ethiopian (n = 264)	Black (n = 532)	White (n = 546)
Prenatal care onset, no. (%)			
1st trimester	147 (63.6)	326 (65.2)	413 (77.9)
2nd or 3rd trimester or no care	84 (36.4)	174 (34.8)	117 (22.1)
Unknown	33 (12.5)	32 (6.0)	16 (2.9)
Diabetes, no. (%)			
None	220 (94.4)	431 (99.3)	504 (98.1)
Established diabetes	2 (0.9)	1 (0.2)	1 (0.2)
Gestational diabetes	11 (4.7)	2 (0.5)	9 (1.8)
Unknown	31 (11.7)	98 (18.4)	32 (5.9)
Anemia during pregnancy,[a] no. (%)			
No	132 (93.6)	226 (86.6)	307 (96.8)
Yes	9 (6.4)	35 (13.4)	10 (3.2)
Unknown	123 (46.6)	271 (50.9)	229 (41.9)
Preeclampsia, no. (%)			
No	216 (92.7)	403 (93.1)	483 (94.2)
Yes	17 (7.3)	30 (6.9)	30 (5.8)
Unknown	31 (11.7)	99 (18.6)	33 (6.0)
Chronic hypertension, no. (%)			
No	232 (99.6)	424 (97.7)	506 (98.6)
Yes	1 (0.4)	10 (2.3)	7 (1.4)
Unknown	31 (11.7)	98 (18.4)	33 (6.0)
Birthweight, no. (%)			
<1499 g	5 (1.9)	12 (2.3)	2 (0.4)
1500–2499 g	9 (3.4)	28 (5.3)	23 (4.2)
2500–3999 g	196 (74.2)	470 (88.4)	429 (78.6)
4000+ g	54 (20.5)	22 (4.1)	92 (16.9)
Unknown	0 (0.0)	0 (0.0)	0 (0.0)
Mean birthweight, g	3486	3176	3492

[a]Anemia information was available on the birth certificate in 1989 through 1991 only.

was due to a large number of Black births near the upper age limit of this category.

The favorable pregnancy outcomes of Ethiopian women relative to US-born Blacks may be due to several factors. The US-born Black and Ethiopian populations may differ in their responses to physical and emotional stress, and pregnant Ethiopian women may receive more psychological and emotional support than US-born Black women. Hispanic women in the United States have been shown to have pregnancy outcomes more like those of Whites than those of Blacks, attributed in part to supportive communities that may help buffer psychological stress and deter harmful behavioral practices such as smoking and alcohol and drug use.[11] This mechanism may be operating in the Ethiopian community as well, as evidenced by the high marriage and low smoking rates among our Ethiopian subjects. There may also be nutritional differences between Ethiopian and US-born Black women before and during pregnancy. This hypothesis is also supported by our data: the frequency of anemia during pregnancy among Ethiopians was half that of US-born Black women. Finally, genetic differences between Ethiopians and more ethnically heterogeneous US-born Black women may account for some of the difference in birthweight distributions.

Future investigation of the Ethiopian community structure, as well as dietetic and behavioral practices during pregnancy, is warranted to illuminate the reasons for the favorable pregnancy outcomes of Ethiopian women. Such research will promote identification of additional modifiable factors applicable to decreasing low birthweight rates among US-born Black women. □

Acknowledgments

We would like to thank Robert Davis, MD, and Michelle Williams, PhD, for their thoughtful review of this manuscript and guiding comments over the course of this project.

References

1. Cabral H, Fried LE, Levenson S, Amaro H, Zuckerman B. Foreign-born and U.S.-born black women: Differences in health behaviors and birth outcomes. *Am J Public Health.* 1990;80:70–72.
2. Centers for Disease Control. Ethnic variation and maternal characteristics among blacks—Massachusetts, 1987 and 1988. *MMWR.* 1991;40:403–411.
3. Valanis BM, Rush D. A partial explanation of superior birth weights among foreign-born women. *Soc Biol.* 1979;26:198–210.
4. Kleinman JC, Fingerhut LA, Prager K. Differences in infant mortality by race, nativity status, and other maternal characteristics. *Am J Dis Child.* 1991;145:194–199.
5. Doucet H, Baumgarten M, Infante-Rivard C. Risk of low birthweight and prematurity among foreign-born mothers. *Can J Public Health.* 1992;83:192–195.
6. Parenti DM, Lucas D, Lee A, Hollenkamp R. Health status of Ethiopian refugees in the United States. *Am J Public Health.* 1987;77:1542–1543.
7. Chaos in the Horn of Africa. *Refugee Rep.* 1991;12:8–11.
8. Ethiopian refugee resettlement in the US. *Refugee Rep.* 1991;12:11–15.
9. Ruiz HA. *Beyond the Headlines: Refugees in the Horn of Africa.* Washington, DC: US Committee for Refugees; 1988.
10. *World Refugee Survey.* Washington, DC: US Committee for Refugees; 1991.
11. James SA. Racial and ethnic differences in infant mortality and low birthweight: a psychosocial critique. *Ann Epidemiol.* 1993; 3:130–136.

The Risk of Pregnancy-Induced Hypertension: Black and White Differences in a Military Population

Debra E. Irwin, PhD, David A. Savitz, PhD, Irva Hertz-Picciotto, PhD, and Kenneth A. St. André, MD, MPH

ABSTRACT

The relationship between race and risk of pregnancy-induced hypertension was investigated in a cohort of active-duty military women who gave birth during the period 1987 through 1989. Cases were identified through hospital discharge diagnoses and included transient gestational hypertension, pre-eclampsia, eclampsia, and unspecified hypertension complicating pregnancy. Multivariate analysis showed nulliparous Black women to be at a slightly increased risk for all pregnancy-induced hypertension (risk ratio [RR] = 1.2) and for pre-eclampsia (RR = 1.3) compared with nulliparous White women. Black parous women were found to have a slightly reduced risk of all pregnancy-induced hypertension (RR = 0.77) and pre-eclampsia (RR = 0.38) compared with White parous women. (*Am J Public Health.* 1994;84:1508–1510)

Introduction

Fetal death, low birthweight, and neonatal mortality rates for Blacks are about twice as high as those for Whites,[1] owing in part to inadequate access to gynecological and prenatal services and to life-style factors associated with low socioeconomic status.[1] Some studies report a higher incidence of pregnancy-induced hypertension among Black women than among White women,[2–4] possibly owing to a predominance of socioeconomic and life-style factors potentially related to the disease and to a higher prevalence of undiagnosed preexisting hypertension. Other studies have not found an increased risk among Blacks.[5–8] The present cohort study examines the relationship between race and risk for pregnancy-induced hypertension among active-duty women in the US Navy and Marine Corps, among whom the health, education, and socioeconomic differences between Black and White populations are significantly reduced and the barriers to medical care have been minimized.

Methods

All active-duty US Navy and Marine Corps personnel who were admitted to all military hospitals for a singleton infant delivery between October 1, 1987, and September 30, 1989, were initially included in the cohort (n = 8375). After those with preexisting hypertension recorded at hospital discharge (82 Whites, 24 Blacks, 10 of other races) were excluded, a total of 8259 women were eligible for the study.

Individual hospital discharge records were obtained from the Retrospective Case-Mix Analysis System, a component database of the Defense Medical Information System, which contains hospital discharge summary data for all hospitalizations from Department of Defense–operated hospitals. Women identified in the cohort were linked on social security number to demographic database files at the Defense Manpower Data Center. More than 89% (n = 7364) of the women who were admitted for infant delivery between the stipulated dates were successfully linked. Women who were and were not linked did not differ with respect to age, rank, branch of service, parity, type of delivery, and pregnancy-induced hypertension status. The 377 women who were neither Black nor White were excluded, leaving 6987 women in the analysis.

Women were classified as having pregnancy-induced hypertension if one of their hospital discharge diagnoses was for that disease as defined by the *International Classification of Diseases,* ninth revision (ICD-9)[9]: transient gestational hypertension (code 642.3), pre-eclampsia (642.4 and 642.5), and all pregnancy-induced hypertension (included codes for transient gestational hypertension, pre-eclampsia, eclampsia [642.6], and unspecified hypertension complicating pregnancy [642.9]). Women whose discharge diagnosis indicated infant delivery but did not indicate any of the above pregnancy-induced hypertension diagnoses were considered nondiseased. Crude risks and risk ratios (RRs) were calculated, and adjusted odds ratios (ORs) were obtained by multivariate logistic regression.[10]

Results

Comparison of the sociodemographic and reproductive characteristics of Blacks and Whites in this study indicates that White women were notably more likely than Black women to be married (75% vs 56%) and slightly more likely to be among the officer ranks (9% vs 2%), to be over age 35 (3.4% vs 1.6%), and to have

Debra E. Irwin, David A. Savitz, and Irva Hertz-Picciotto are with the School of Public Health at the University of North Carolina in Chapel Hill. Kenneth A. St. André is with the Office of the Assistant Secretary of Defense (Health Affairs), Department of Defense, Washington, DC.

Requests for reprints should be sent to Debra E. Irwin, PhD, 3030 Cornwallis Rd, Research Triangle Park, NC 27709.

This paper was accepted December 8, 1993.

education past high school (21% vs 17%). Equal proportions of Whites and Blacks were nulliparas, and the proportions diagnosed with pregnancy-induced hypertension were similar. Nulliparas were at increased risk of all pregnancy-induced hypertension (RR = 2.2; 95% confidence interval [CI] = 1.8, 2.7), transient gestational hypertension (RR = 2.0; 95% CI = 1.5, 2.7), and pre-eclampsia (RR = 2.8; 95% CI = 2.0, 3.9) compared with parous women, but there was little association between age, race, marital status, or educational level and risk of these three conditions.

Nulliparous Black women had modestly increased risk ratios for all pregnancy-induced hypertension and pre-eclampsia but not for transient gestational hypertension (Table 1). Among parous women, Blacks were at reduced risk for pre-eclampsia compared with Whites, although both races appear to be at equal risk for the development of all pregnancy-induced hypertension and transient gestational hypertension (Table 2).

TABLE 1—Multivariate[a] Analysis of Demographic Characteristics and Risks for Pregnancy-Induced Hypertension, Transient Gestational Hypertension, and Pre-Eclampsia among Nulliparous Women

	All Pregnancy-Induced Hypertension			Transient Gestational Hypertension			Pre-Eclampsia		
	Crude Risk	OR[a]	95% CI	Crude Risk	OR[a]	95% CI	Crude Risk	OR[a]	95% CI
Race									
White	10.0	1.0	. . .	5.2	1.0	. . .	4.4	1.0	. . .
Black	11.7	1.2	0.97, 1.5	4.8	0.96	0.68, 1.4	5.7	1.3	0.97, 1.7
Service									
Navy	9.5	1.0	. . .	3.7	1.0	. . .	5.2	1.0	. . .
Marines	11.0	1.2	0.92, 1.5	5.4	1.5	0.99, 2.2	4.8	0.95	0.67, 1.4
Military rank									
Officer	12.6	1.0	. . .	6.7	1.0	. . .	4.7	1.0	. . .
Enlisted	10.6	0.82	0.56, 1.2	5.0	0.79	0.49, 1.3	4.9	1.1	0.58, 1.9
Age, y									
17–34	10.6	1.0	. . .	5.0	1.0	. . .	4.9	1.0	. . .
≥35	13.8	1.2	0.62, 2.4	10.3	2.0	0.95, 2.2	3.5	0.54	0.13, 2.3

Note. OR = odds ratio; CI = confidence interval.
[a]Multiple logistic regression model controlling for the effects of race, age, military rank, and military service.

TABLE 2—Multivariate[a] Analysis of Demographic Characteristics and Risks for Pregnancy-Induced Hypertension, Transient Gestational Hypertension, and Pre-Eclampsia among Parous Women

	All Pregnancy-Induced Hypertension			Transient Gestational Hypertension			Pre-Eclampsia		
	Crude Risk	OR[a]	95% CI	Crude Risk	OR[a]	95% CI	Crude Risk	OR[a]	95% CI
Race									
White	5.0	1.0	. . .	2.5	1.0	. . .	2.1	1.0	. . .
Black	4.0	0.82	0.51, 1.3	2.7	1.1	0.66, 2.1	0.8	0.37	0.15, 0.95
Service									
Navy	4.0	1.0	. . .	1.5	1.0	. . .	2.0	1.0	. . .
Marines	4.9	1.2	0.67, 2.1	2.8	2.1	0.83, 5.2	1.8	0.73	0.33, 1.6
Military rank									
Officer	6.1	1.0	. . .	4.0	1.0	. . .	2.0	1.0	. . .
Enlisted	4.6	0.89	0.45, 1.8	2.4	0.63	0.27, 1.5	1.8	1.2	0.36, 3.6
Age, y									
17–34	4.5	1.0	. . .	2.4	1.0	. . .	1.7	1.0	. . .
≥35	9.1	1.7	0.80, 3.6	5.3	1.3	0.33, 3.5	3.0	2.0	0.13, 2.3

Note. OR = odds ratio; CI = confidence interval.
[a]Multiple logistic regression model controlling for the effects of race, age, military rank, and military service.

Discussion

The current study found a slightly increased risk for pre-eclampsia among nulliparous Black women compared with nulliparous White women in contrast to a fairly strong inverse association for pre-eclampsia among parous Black women compared with parous White women. Eskenazi et al.[2] found nulliparous Blacks to have a dramatically increased risk for pre-eclampsia compared with their White counterparts (OR = 12.3; 95% CI = 1.6, 100.8), whereas parous Blacks were found to have a reduced risk for pre-eclampsia compared with parous Whites (OR = 0.68; 95% CI = 0.13, 3.5). The current study's findings for nulliparous and parous Black women are in the same direction but of different magnitude than Eskenazi et al.'s[2] results. Other studies have reported conflicting findings.[7]

Evaluation of these data requires consideration of the accuracy of the diagnoses recorded in the database. As a regular part of the hospital quality assurance process, the military services have instituted the Civilian External Peer Review program, in which nonmilitary physicians evaluate the accuracy of patient care and diagnoses. This survey showed that inpatient hospital care for pre-eclampsia was found to be more than 97% compliant with diagnostic and case management criteria,[11] limited to evidence concerning false-positive diagnosis but not false-negative diagnosis.

The unique nature of the active-duty military population reduces some of the differences in nutrition, physical activity, and quality of health care that may have affected racial comparisons in previous studies. The preenlistment experiences of Black and White women undoubtedly differ, but at least for their time in the military, racial groups are constrained to be more similar than they would be as civilians: eligibility requirements for enlistment include a high school education or its equivalent; physical standards for initial enlistment prevent individuals with any significant chronic illness or disability from entering on active duty; all candi-

dates for active duty are tested for evidence of illicit drug use on induction into the military and periodically thereafter; and twice each year, individuals must successfully pass a physical fitness test that includes a test of aerobic capacity, flexibility, strength, and body fat content. These standards produce a relatively homogeneous active-duty population that is moderately well educated, physically fit, and essentially disease free. Medical care for preventive and therapeutic services is free of charge and readily available to active-duty personnel. Time for prenatal medical examinations is provided as part of normal working hours, and women are strongly encouraged to obtain early and consistent prenatal care. The selectivity of this population is evident in its lower rates of low birthweight risk (Blacks 6%, Whites 4%) and reduced racial differences in low birthweight risk compared with the general US population (Blacks 13%, Whites 6%).[1]

It is possible that the similar rates of pregnancy-induced hypertension between the racial subgroups and the decreased risks of pre-eclampsia for multiparous Black women found in this study are due to unmeasured confounders. For example, smoking is a strong protective factor for pregnancy-induced hypertension that was not available for analysis and thus was not controlled. A similar study among pregnant women in the military found minimal differences in smoking rates between Blacks and Whites (18% vs 20% smokers) (unpublished data). An alternative explanation for the results found for multiparous women may be differences in the characteristics of Blacks and Whites who stay in the military after giving birth to their first child.

This study suggests that previously reported elevated risks of pregnancy-induced hypertension in Black as compared with White women may be partially explained by poor access to medical care or by a higher prevalence of chronic disease and generally poorer physical fitness. The apparent interaction between race and parity in the etiology of pre-eclampsia warrants further evaluation. Although the stringent selection criteria for entry into the military limits the generalizability of these results, active-duty military personnel serve as a useful population for examining the risks of adverse pregnancy outcome without the potential biasing effects of differential access to medical care. □

Acknowledgments

This paper was presented at the 120th Annual Meeting of the American Public Health Association; November 8–12, 1992; Washington, DC.

We would like to thank Deborah Eitelberg for her assistance in obtaining the demographic database files at the Defense Manpower Data Center.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication PHS 91–50212.
2. Eskenazi B, Fenster L, Sidney S. A multivariate analysis of risk factors for pre-eclampsia. *JAMA.* 1991;266:237–241.
3. Hendricks CH. Delivery patterns and reproductive efficiency among groups of differing socioeconomic status and ethnic origin. *Am J Obstet Gynecol.* 1967;97:608–624.
4. Hulsey TC, Levkoff AH, Alexander GR, Thomas M. Differences in Black and White infant birth weights: the role of maternal demographic factors and medical complications of pregnancy. *South Med J.* 1991;84:443–446.
5. Chesley LC. History and epidemiology of preeclampsia-eclampsia. *Clin Obstet Gynecol.* 1984;27:801–820.
6. Niswander KR, Gordon M, *The Collaborative Perinatal Study of the National Institute of Neurological Diseases and Stroke: The Women and Their Pregnancies.* Philadelphia, Pa: WB Saunders Co; 1972.
7. Savitz DA, Zhang J. Pregnancy-induced hypertension in North Carolina, 1988 and 1989. *Am J Public Health.* 1992;82:675–679.
8. Saftlas AF, Olson DR, Franks AL, Atrash HK, Pokras R. Epidemiology of pre-eclampsia and eclampsia in the United States, 1979–1986. *Am J Obstet Gynecol.* 1990;163:460–465.
9. *International Classification of Diseases.* 9th Revision, Clinical Modification. 3rd ed. Hyattsville, Md: National Center for Health Statistics; 1989. DHHS publication PHS 89–1260.
10. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research Principles and Quantitative Methods.* New York, NY: Van Nostrand Reinhold Co; 1982:419–491.
11. Buck AS, Martin ED, Mazzuchi JF, Merry M, Mendez E. The Department of Defense Civilian External Peer Review Program: an interim report. *Milit Med.* 1992; 157:40–46.

Call for Papers on Substance Use, Homelessness, and Refugees

Researchers are invited to submit papers for review and possible inclusion in a special issue of the *International Journal of the Addictions* on the subject of "Substance Use, Homelessness, and Refugees: An International Overview." Manuscripts should be postmarked by *December 10, 1994,* and submitted in triplicate to Dr Timothy Johnson, Associate Director, Survey Research Laboratory, University of Illinois, PO Box 6905, Chicago, IL 60680; tel (312) 996-5310.

Nonclient Factors in the Reporting of Prenatal Psychosocial Risk Assessments

Deborah Schild Wilkinson, MSW, MPH, Carol C. Korenbrot, PhD, and Elena Fuentes-Afflick, MD, MPH

ABSTRACT

Psychosocial and nutrition risk assessments at 27 randomly selected sites of a Medicaid-funded enhanced prenatal program were examined for completeness and uniformity of reporting. The recording of psychosocial data, but not of nutrition data, was found to be associated with a substantial number of nonclient factors, including geographical location and abstractor, assessor, and service provider characteristics. The variability in recording of psychosocial problems reflects the numerous ways that assessments are performed and the subsequent difficulty in interpreting the practice of psychosocial assessment. A uniform system of defining and classifying risks is needed to improve both service delivery and service evaluation. (*Am J Public Health.* 1994;84:1511–1514)

Introduction

Psychosocial services are a growing part of the content of prenatal care for low-income women.[1] The Comprehensive Perinatal Services Program of the California Department of Health Services provides such enhanced prenatal and postnatal care services to Medicaid-eligible pregnant women. As the first step of an attempt to evaluate the associations of psychosocial services with outcomes for women in the program, we investigated whether the documentation of the assessment of five psychosocial problems and one strength was complete and uniform. We identified several nonclient factors associated with the recording of psychosocial risks that were not found in the recording of nutrition risks of the same women. We then examined how the association of reporting of each problem and strength with its region was changed by adjusting for three nonclient factors. The findings have important implications for psychosocial assessment and practice and for the evaluation of their impact on health outcomes.

Methods

The study sample, drawn from a stratified random sample of Comprehensive Perinatal Services Program providers (n = 27) certified by the state Department of Health Services, consisted of 3459 Medicaid-eligible women giving birth from July 1989 to December 1990 (C. Korenbrot et al., unpublished data, 1994).[2] Selection of the sites had been stratified for two metropolitan (Los Angeles and the San Francisco Bay Area) and two nonmetropolitan (Sacramento and areas surrounding Fresno) regions of the state.

Potential psychosocial and nutrition risks were assessed by site personnel who conducted interviews with each pregnant woman. Assessment information was documented on site-specific, state-approved forms.

Any health professional at a Comprehensive Perinatal Services Program site may provide any or all of the support services. A generalist position called "comprehensive perinatal health worker" was created for the program. Such a worker must be more than 18 years of age, have a high school diploma or its equivalent, and have at least 1 year of full-time paid practical experience providing services in a perinatal setting. For this study, the psychosocial assessors were grouped into five categories: social workers (BSW, MSW), nurses (LVN, RN, BSN, PHN, FNP), those with other bachelor's degrees, those with other master's degrees, and comprehensive perinatal health workers. There were four groups of nutrition assessors: registered dietitians, nurses, those with other degrees, and comprehensive perinatal health workers.

Five abstractors were accredited records technicians who collected the data from all but one site. A sixth abstractor, a social worker, collected the data from one site only. Charts from each site were abstracted by only one abstractor. Abstractor training took place in two daylong sessions that included having all technicians abstract the same chart at the end of each session. The procedure was repeated three times during the data collection period to ensure continued reliability. Differences in abstraction of particular items were noted, and feedback was given individually and in group meetings to remove the differences. All abstractors were in continual telephone contact with a single project director who resolved difficulties that occurred with abstracting

Deborah Schild Wilkinson is with the School of Social Welfare, University of California, Berkeley. Carol C. Korenbrot is with the Institute for Health Policy Studies and the Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco. Elena Fuentes-Afflick is with the Institute for Health Policy Studies, and the Department of Pediatrics, University of California, San Francisco.

Requests for reprints should be sent to Deborah Schild Wilkinson, MSW, MPH, School of Social Welfare, University of California at Berkeley, 120 Haviland Hall, Berkeley, CA 94720.

This paper was accepted March 16, 1994.

TABLE 1—Proportion of Women with Reported Psychosocial Problem or Strength, by Region

Problem or Strength	Region: Metropolitan (n = 1496), %	Region: Nonmetropolitan (n = 1963), %
Lack of social support	17.3	24.6[a]
Supportive partner	66.9	75.9[a]
Unplanned pregnancy	35.3	57.9[a]
Financial problems	58.4	68.7[a]
Parenting problems	18.9	46.9[a]
Psychological problems	33.2	46.2[a]

[a]Significantly different at the $P < .0001$ level.

data for individual cases. One abstractor was noted to record data in more detail than the others, including margin notes. That abstractor was cautioned against inferring data from the chart notes.

To abstract the psychosocial data as uniformly as possible from the 18 assessment forms used at the 27 sites, a list was compiled associating 122 different phrases used on the forms with 72 different codes. The codes were combined into 27 theoretically consistent risks, each composed of one or more of the codes. Each code was used in only one risk. The risks were constructed as dichotomous variables such that if one or more of the codes was present, the value was 1; otherwise, the value was 0. Five problem areas were included in the analysis: (1) lack of social support, (2) unplanned pregnancy, (3) financial problems, (4) potential and existing parenting problems, and (5) psychological problems. Unplanned pregnancy was composed of a single code, while other risks had several. Lack of social support, for example, had four codes composed of six specific phrases: minimal or no emotional support; lacks social contacts; feels alone or lonely; lives alone (as a risk); multiple partners, numerous unstable relationships; and other social support problems. Other psychosocial data concerning client strengths were collected. One strength (support from husband or partner and/or father of baby) was included in the analysis.

Nine nutrition risks could be coded on the abstraction forms if noted by an assessor in the chart: (1) anemia, (2) chronic disease, (3) eating disorder, (4) high weight gain, (5) low weight gain, (6) inadequate diet, (7) obese before pregnancy, (8) short pregnancy interval, and (9) underweight before pregnancy.

Correlations between the psychosocial and nutrition risks and each of the reporting variables—metropolitan or nonmetropolitan area, data abstractor, educational credential of the risk assessor, and provider type—were measured. Forced hierarchical multiple logistic regression models were then constructed for each dependent psychosocial variable. In the first model, population density (nonmetropolitan) was the sole explanatory variable. In the second hierarchy step, the changes in log-likelihood relative to the first model of models adding only abstractor, assessor's credentials, or provider type were examined; the model with the largest change (abstractor) was selected to be the base of the following models to which each of the two remaining variable groups were added in turn. The final model included all of the variable groups.

Results

Women in the metropolitan areas of the study were unexpectedly found to be significantly less likely to have psychosocial risks than women in nonmetropolitan areas ($P < .0001$; Table 1). We examined the correlation of nonclient factors with the reporting of these psychosocial risks (Table 2) and examined similar correlations for nutrition risks (Table 3). The relative size and significance of the unadjusted correlation of each of the nonclient factors were greater for psychosocial risks (Table 2) than for nutrition risks (Table 3). Thirty-three percent of the correlations for psychosocial risks were greater in magnitude than .10 and were significant at the $P < .001$ level. In contrast, only 10% of the correlations for nutrition risks reached .10, and nearly half of these high correlations involved a single risk: inadequate diet.

Adjustment for abstractor alone removed the association with region for the two support variables (Table 4). Adjustment for all three explanatory variables removed the association of unplanned pregnancy with region. Adjustment for nonclient factors reversed the regional effect for financial problems, resulting in the finding that women in metropolitan regions were more likely, not less likely, to have financial problems. Finally, for two psychosocial risks, potential or existing parenting problems and psychological problems, the regional effect remained regardless of the adjustment model used.

Discussion

In its guidelines for prenatal care, the Public Health Service's Expert Panel[1] has set out a challenge to include in prenatal care a new emphasis on the psychosocial dimensions of care. However, evidence that psychosocial services in practice are associated with the improved health and well-being of pregnant women needs further documentation.[3] The findings of our study demonstrate how the lack of a uniform method of defining and classifying problems makes it impossible to evaluate appropriately the associated impact of psychosocial services in practice. If client problems cannot be systematically classified, then reductions in problems and associated changes in health status for women with problems who receive interventions cannot be evaluated.

Any classification system for psychosocial risks in health care settings must have clear, consistent phraseology and address both the objective and subjective factors related to the problems.[3-8] The system must be practical for use in the field and lead to appropriate interventions as well, and it must be flexible enough to accommodate the cultural differences between populations served. The use of commonly endorsed systems makes the effectiveness of psychosocial services testable. □

Acknowledgments

This research was funded by the Bureau of Maternal and Child Health, Public Health Service, US Department of Health and Human Services (grant MCJ-067-010).

Dr Fuentes-Afflick received fellowship support from the Hewlett and Pew Foundations during the course of this study.

TABLE 2—Correlation Matrix of Reported Psychosocial Problems with Nonclient Factors

Problem or Strength	Region: Metropolitan	Abstractor						Assessor					Provider				
		1	2	3	4	5	6	Social Worker	Nurse	BA Degree	Other Degree	CPHW	MD Office	Community Clinic	Health Department	Public Hospital	Private Hospital
Lack of social support	*−.113*	.055	.052	−.030	*−.101*	.066	−.074	−.081	.076	−.051	−.030	.089	−.037	.009	.022	.016	−.009
Supportive partner	−.091	.020	−.006	.044	*−.119*	.073	−.017	−.047	.028	.028	−.060	.041	−.004	.016	.014	−.029	−.006
Unplanned pregnancy	*−.218*	.041	.078	*−.152*	*−.166*	*.268*	*−.117*	*−.151*	*.173*	*−.107*	−.007	*.109*	*−.109*	.077	−.042	.071	.003
Financial problems	*−.179*	*−.180*	−.064	*.287*	*−.274*	*.239*	.026	*−.211*	*.104*	.076	−.084	*.148*	−.048	.022	−.009	−.004	.036
Parenting problems	*−.341*	*.128*	−.034	*−.118*	*−.304*	*.317*	−.042	*−.178*	*.288*	−.091	−.039	−.004	*.134*	*−.122*	−.072	.016	.068
Psychological problems	*−.124*	−.046	.084	−.016	*−.124*	*.152*	−.070	−.016	.014	−.057	.002	.064	−.027	.058	−.018	−.027	−.001

Note. CPHW = comprehensive perinatal health worker. All absolute values greater than or equal to .100 are italicized, and all italicized correlations are statistically significant at $P < .001$.

TABLE 3—Correlation Matrix of Reported Nutritional Problems with Nonclient Factors

Problem or Strength	Region: Metropolitan	Abstractor						Assessor				Provider				
		1	2	3	4	5	6	RD	Nurse	Other Degree	CPHW	MD Office	Community Clinic	Health Department	Public Hospital	Private Hospital
Anemia	.097	−.021	−.031	.014	.051	.017	.021	−.079	.035	.004	.079	.038	.026	.026	.045	−.090
Chronic disease	−.052	.095	−.041	−.023	−.041	−.039	.055	*.103*	−.042	−.053	−.043	*.134*	−.064	−.041	−.019	−.015
Eating disorder	−.075	.041	*.112*	−.058	−.049	−.005	−.033	*−.110*	.012	*.107*	.019	−.049	.021	.085	−.011	−.033
High weight gain	.018	−.083	−.055	−.001	.040	.053	.053	*.105*	−.015	−.047	−.087	.008	−.085	−.025	.028	.073
Low weight gain	.031	−.033	.035	.019	−.039	.047	−.007	.050	−.062	−.056	.071	−.080	.052	.087	.034	−.067
Inadequate diet	*.112*	−.068	*.179*	*.330*	*−.165*	*−.178*	.089	−.015	.042	−.064	.061	.082	.071	*.158*	.009	*−.226*
Obese before pregnancy	−.058	−.022	−.042	*−.115*	−.010	*.166*	.059	−.022	.088	.003	−.074	.031	−.036	−.029	.052	.005
Short pregnancy interval	−.037	−.024	.033	.02:	−.028	.036	.013	−.040	.009	.033	.009	−.029	.034	−.011	−.042	.073
Underweight before pregnancy	.015	−.098	.051	−.047	.029	.047	.046	.035	−.059	.013	−.002	−.037	.039	.002	−.003	−.012

Note. RD = registered dietitian; CPHW = comprehensive perinatal health worker. All absolute values greater than or equal to .100 are italicized, and all italicized correlations are statistically significant at $P < .001$.

TABLE 4—The Recording of Psychosocial Strength or Problem in Assessment Data: Its Dependence on Nonclient Factors

Psychosocial Strength or Problem	Unadjusted Odds Ratio for Metropolitan (95% CI)	Adjusted Odds Ratio for Metropolitan (95% CI)	Variables Included in Adjusted Model
Lack of social support	0.584 (0.494, 0.690)	0.929 (0.749, 1.152)	Abstractor only
Supportive partner	0.661 (0.571, 0.766)	0.842 (0.685, 1.034)	Abstractor only
Unplanned pregnancy	0.412 (0.352, 0.472)	0.786 (0.615, 1.010)	Abstractor Assessor credential Provider type
Financial problems	0.456 (0.397, 0.524)	2.073 (1.505, 2.856)	Abstractor Assessor credential Provider type
Parenting problems	0.210 (0.179, 0.245)	0.277 (0.220, 0.350)	Abstractor Assessor credential Provider type
Psychological problems	0.547 (0.477, 0.628)	0.728 (0.572, 0.925)	Abstractor Assessor credential Provider type

We wish to thank Ayesha E. Gill, PhD, of the Institute for Health Policy Studies, University of California, San Francisco; Athole Lennie, DrPH, of the Comprehensive Perinatal Services Program; and Xavier Castorena of the California Department of Health Services for their assistance, comments, and support.

References

1. *Caring for Our Future: The Content of Prenatal Care.* Washington, DC: US Dept of Health and Human Services; 1989.
2. Korenbrot C, Clayson Z, Gill A, Patterson E. Evaluation of the statewide implementation of enhanced perinatal services as Medicaid benefits. *Public Health Rep.* In press.
3. Thompson J. Maternal stress, anxiety, and social support during pregnancy: possible directions for prenatal intervention. In: Merkatz IR, Thompson JE, eds. *New Perspectives on Prenatal Care.* New York, NY: Elsevier Science; 1990:319–335.
4. Henk M. Developing codes and classifications for social workers in health care. In: Gitterman A, Black RB, Stein F, eds. *Public Health Social Work in Maternal and Child Health: A Forward Plan.* New York, NY: Columbia University; 1985:125–131.
5. Institute of Medicine. *Nutrition during Pregnancy.* Washington, DC: National Academy Press; 1990.
6. Guendelman S. Screening for psychosocial risk. In: Morton CJ, Balassone ML, Guendelman S, eds. *Preventing Low Birthweight and Infant Mortality: Programmatic Issues for Public Health Social Workers.* Berkeley, Calif: Maternal and Child Health Program, School of Public Health, University of California; 1987:85–88.
7. Coulton CJ. *Social Work Quality Assurance Programs: A Comparative Analysis.* Washington, DC: National Association of Social Workers; 1979.
8. Norton PG, Stewart M, Tudiver F, Bass MJ, Dunn EV, eds. *Primary Care Research: Traditional Innovative Approaches.* Newbury Park, Calif: Sage Publications; 1991.

A2623-1

American Journal of Public Health

October 1994, Vol. 84, No. 10 Established 1911

ED KIANKA
HEALTH REPORTER
CONGRESSIONAL INFORMATION SVC
4520 EAST-WEST HIGHWAY 9375
BETHESDA MD 20814

American Public Health Association

Featuring Medical Care

Editorials

1566 Health Care Reform and Health Outcomes *L. S. Robertson*
1567 Toward a More Population-Based Approach to Immunization: Fostering Private- and Public-Sector Collaboration *R. H. Bernier*

Commentary

1569 Developing Leaders vs Training Administrators in the Health Services *M. W. Legnini*

Public Health Policy Forum

1573 Changes in Positions of Authority Held by US Physicians: A Fresh Look at Existing Data *J. S. Osberg*

Articles

VARIOUS TOPICS

1576 Estimates of Population Smoking Prevalence: Self- vs Proxy Reports of Smoking Status *E. A. Gilpin, J. P. Pierce, S. W. Cavin, C. C. Berry, N. J. Evans, M. Johnson, and D. G. Bal*
1580 Job Strain and Health-Related Quality of Life in a National Sample *D. J. Lerner, S. Levine, S. Malspeis, and R. B. D'Agostino*
1586 Alcohol and Drug Disorders among Physically Abusive and Neglectful Parents in a Community-Based Sample *K. Kelleher, M. Chaffin, J. Hollenberg, and E. Fischer*

MEDICAL CARE

1591 Consequences of the North American Free Trade Agreement for Health Services: A Perspective from Mexico *J. Frenk, O. Gómez-Dantés, C. Cruz, F. Chacón, P. Hernández, and P. Freeman*
1598 Stage of Cancer at Diagnosis for Medicare HMO and Fee-for-Service Enrollees *G. F. Riley, A. L. Potosky, J. D. Lubitz, and M. L. Brown*
1605 Acutely Injured Patients with Trauma in Massachusetts: Differences in Care and Mortality, by Insurance Status *J. S. Haas and L. Goldman*
1609 Comparing Hospitals That Perform Coronary Artery Bypass Surgery: The Effect of Outcome Measures and Data Sources *A. J. Hartz and E. M. Kuhn*
1615 Rates, Patterns, Causes, and Costs of Hospitalization of Nursing Home Residents: A Population-Based Study *W. H. Barker, J. G. Zimmer, W. J. Hall, B. C. Ruff, C. B. Freundlich, and G. M. Eggert*
1621 Risk Factors for Delayed Immunization among Children in an HMO *T. A. Lieu, S. B. Black, P. Ray, M. Chellino, H. R. Shinefield, and N. E. Adler*
1626 Presumptive Eligibility for Pregnant Medicaid Enrollees: Its Effects on Prenatal Care and Perinatal Outcome *J. M. Piper, E. F. Mitchel, Jr, and W. A. Ray*
1631 The Danger of Applying Uniform Clinical Policies across Populations: The Case of Breast Cancer in American Indians *P. A. Nutting, B. N. Calonge, D. C. Iverson, and L. A. Green*

Public Health Briefs

VARIOUS TOPICS

1637 HIV Infection among Women in Prison: An Assessment of Risk Factors Using a Nonnominal Methodology *C. A. Hankins, S. Gendron, M. A. Handley, C. Richard, M. T. Lai Tung, and M. O'Shaughnessy*

(continued on page 1553)

Contents

American Journal of Public Health

The Journal's one-page author guide, "What *AJPH* Authors Should Know," is printed in each issue and is available from the Editorial Office.

American Public Health Association

October 1994, Vol. 84, No. 10 ISSN: 0090-0036 Established 1911

1640 Risk Factors for Acute Non-A, Non-B Hepatitis and Their Relationship to Antibodies for Hepatitis C Virus: A Case–Control Study *A. Mele, L. Saglíocca, G. Manzillo, F. Converti, P. Amoroso, M. A. Stazi, L. Ferrigno, M. Rapicetta, E. Franco, B. Adamo, F. Palumbo, C. Sbreglia, A. Panà, P. Pasquini, and the SEIEVA Collaborating Group*

MEDICAL CARE

1644 CDC's Consensus Set of Health Status Indicators: Monitoring and Prioritization by State Health Departments *S. L. Zucconi and C. A. Carson*

1646 Medicare-Certified Home Health Services: National and Regional Supply in the 1980s *C. C. Scalzi, J. S. Zinn, M. J. Guilfoyle, and S. T. Perdue*

1649 Gynecologists' Sex, Clinical Beliefs, and Hysterectomy Rates *N. A. Bickell, J. A. Earp, J. M. Garrett, and A. T. Evans*

1653 Hysterectomy Use: The Correspondence between Self-Reports and Hospital Records *K. M. Brett and J. H. Madans*

1655 Recruitment Activities and Sociodemographic Factors That Predict Attendance at a Mammographic Screening Program *S. F. Hurley, R. M. Huggins, D. J. Jolley, and D. Reading*

1658 The Accessibility of Drug Treatment for Pregnant Women: A Survey of Programs in Five Cities *V. Breitbart, W. Chavkin, and P. H. Wise*

1662 National and State Spending on Specialty Alcoholism Treatment: 1979 and 1989 *J. H. Huber, G. C. Pope, and D. A. Dayhoff*

1666 Monitoring Childhood Immunizations: A Canadian Approach *J. D. Roberts, L. A. Poffenroth, L. L. Roos, J. D. Bebchuk, and A. O. Carter*

1669 Patient Adherence to Guidelines for Diabetes Eye Care: Results from the Diabetic Eye Disease Follow-Up Study *J. C. Will, R. R. German, E. Schuman, S. Michael, D. M. Kurth, and L. Deeb*

1672 The Risk of Hospitalization for Ischemic Heart Disease among Asian Americans in Northern California *A. L. Klatsky, I. Tekawa, M. A. Armstrong, and S. Sidney*

1675 The Effects of Denturism: New Zealand Dentists' Response to Competition *N. J. Devlin*

Public Health Then and Now

1678 The Medical Care Programs of the Farm Security Administration, 1932 through 1947: A Rehearsal for National Health Insurance? *M. R. Grey*

Notes from the Field

1688 Community-Based Services for Pregnant Substance-Using Women *S. V. LaFrance, J. Mitchell, K. Damus, C. Driver, G. Roman, E. Graham, and L. Schwartz*

1689 Going Smoke-Free in the 1990s: Lessons Learned at a Teaching Hospital *S. E. Radecki and S. A. Brunton*

1691 The Central Harlem Playground Injury Prevention Project: A Model for Change *D. Laraque, B. Barlow, L. Davidson, and C. Welborn*

Other Journal Departments

1554 APHA Masthead

1557 Advertisers' Index

1558 What *AJPH* Authors Should Know

1566 Journal Masthead

1693 Letters to the Editor

1698 Book Corner

1703 Annual Meeting Housing Information

1705 Job Opportunities

American Journal of Public Health

American Public Health Association

The American Journal of Public Health, published monthly, is the official Journal of the American Public Health Association, 1015 15th St, NW, Washington, DC 20005; (202) 789-5600; TDD (202) 789-5673. APHA is a nongovernmental professional society founded in 1872, representing all disciplines and specialties in public health.

APHA OFFICERS

Eugene Feingold, PhD, JD, *President*

Helen Rodriguez-Trias, MD, *Immediate Past President*

Caswell A. Evans, Jr, DDS, MPH, *President-Elect*

Leonard M. Schuman, MD, *Vice-President (USA)*

John E. F. Hastings, MD, MPH, *Vice-President (Canada)*

Oscar Feo, MD, MSc, *Vice-President (Latin America)*

Jay H. Glasser, PhD, *Treasurer*

Allan G. Rosenfield, MD, *Chair, Executive Board*

Shirley F. Randolph, MSPH, *Speaker, Governing Council*

Fernando M. Treviño, PhD, MPH, *Executive Director*

EXECUTIVE BOARD

Allan G. Rosenfield, MD (1994), *Chair*

Carol Easley Allan, PhD, RN (1997)

Caswell A. Evans, Jr, DDS, MPH, *President-Elect*

Patricia P. Evans, MPH (1994)

Eugene Feingold, PhD, JD, *President*

Jay H. Glasser, PhD, *Treasurer*

Alan R. Hinman, MD, MPH (1995)

Margo D. Kerrigan, MPH (1995)

Joanne Lukomnik, MD, MPH (1996) *Vice-Chair*

Steven Uranga McKane, DMD, MPH (1995)

Carlos W. Molina, EdD (1997)

Linda Rae Murray, MD, MPH (1996)

Hila Richardson, DrPH, MPH (1996)

Helen Rodriguez-Trias, MD, *Immediate Past President*

Reed V. Tuckson, MD (1994)

Faye L. Wong, MPH, RD (1997)

EX OFFICIO

Fernando M. Treviño, PhD, MPH, *Executive Director*

Frances Kissling, *Chair, Action Board*

George E. Pickett, MD, MPH, *Speaker, Program Development Board*

Shirley F. Randolph, MSPH, *Speaker, Governing Council*

The *American Journal of Public Health* (*AJPH*), published monthly, is the official Journal of the American Public Health Association, Inc, 1015 15th St, NW, Washington, DC 20005. Founded in 1872, the American Public Health Association is organized to protect and promote personal and environmental health by exercising leadership in the development and dissemination of health policy. Representing all disciplines and specialties in public health, APHA is the largest public health association in the world, with an aggregate membership of 50 000, including its state and local affiliates.

Manuscripts, Correspondence, and Letters to the Editor: Send to Mervyn Susser, DPH, Editor, *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005.

Information on Manuscript Preparation is published monthly in the Journal. The Journal publishes reports of original research, research methods, program evaluations, analytic reviews, including health policy analysis, and reports for special departments. A paper is considered with the understanding that it or its essential substance neither has been published nor is under consideration elsewhere.

Subscription price, individual, is $100/year US, $140/year international; **institutional,** $160/year US, $200/year international. Single copies $13 US, $14 international (surface delivery), $22 international (air delivery). Prepaid only.

Postal Information: Second class postage paid at Washington, DC, and additional mailing offices. Postmaster: Send address changes to *American Journal of Public Health*, 1015 15th St, NW, Washington, DC 20005. *AJPH* allotted $40; *The Nation's Health* allotted $6 of APHA membership dues. Information on individual membership appears elsewhere in this Journal, or contact Director, Membership Services, at APHA headquarters. Printed in the United States.

Advertising Sales: Maureen Sheridan, Advertising Manager, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5668 or 5600. Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the Journal does not indicate or imply endorsement by the Journal or APHA.

Reprints: Request single reprints of articles published in the Journal from the authors at the address given in the footnote on the first page of published article. For cost information on 100 or more reprints, contact Marilyn Butler, *AJPH* Reprints, APHA, 1015 15th St, NW, Washington, DC 20005; (202) 789-5649.

Back Issues of the Journal: A limited inventory of back issues and volumes of the Journal is available at current subscription rates plus handling charge. All orders under $35 must be prepaid. Vols 1 to present are also available on microfilm.

Copyright © 1994 by the American Public Health Association, Inc. All material subject to this copyright may be photocopied for the noncommercial purposes of scientific or educational advancement. Opinions expressed by authors of articles summarized, quoted, or published in full in this Journal represent the opinions of the authors and do not necessarily reflect the official policy of the American Public Health Association or the institution with which the author(s) is (are) affiliated, unless so specified.

Any report, article, or paper prepared by employees of the US Government as part of their official duties is, under the Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the US Code is not available. However, the Journal format is copyrighted and pages may not be photocopied without permission of the American Public Health Association, except in limited quantities for educational purposes, as stated above.

Copying done for other than personal or internal reference use—such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale—without the expressed permission of the American Public Health Association is prohibited. Requests for special permission or bulk orders should be addressed to APHA headquarters.

The *American Journal of Public Health* is indexed in whole or in part in these data bases: Bibliographic Index of Health Education Periodicals, Biological Abstracts, Biology Digest, BRS/COLLEAGUE, Chemical Abstracts, Cumulative Index to Nursing & Allied Health Literature, ETOH, Environmental Information Center Inc, Environmental Periodicals Bibliography, Excerpta Medica, Grateful Med, Hospital Literature Index, Hospital Management Review, Human Resources Abstracts, Index Medicus (Abridged), MEDLARS, MEDSOC, Nuclear Science Abstracts, Physical Education Index, Public Affairs Information Service, Public Health Reviews, Safety Science Abstracts, Sage Family Studies Abstracts, Science Citation Index, and Social Work Research and Abstracts.

Public Health Policy Forum

Changes in Positions of Authority Held by US Physicians: A Fresh Look at Existing Data

ABSTRACT

Some experts contend that physicians have lost positions of authority in the past 40 years. Others argue the opposite, yet neither side bases its arguments on empirical data. This study examined longitudinal variables measuring authority positions held by physicians. Data on the relative position of physicians in medical schools show that medical doctors held 65.6% of the sampled positions in 1970 vs 72.8% in 1990. Yet, in the wider society and within the nonmedical school portion of the health sector, other data indicate that physicians occupy a smaller proportion of authority positions. (*Am J Public Health.* 1994;84:1573–1575)

J. Scott Osberg, PhD

Introduction

Data on positions of authority held by US physicians are widely available, but researchers have not systematically analyzed them. This paper synthesizes existing data on physicians' positions of authority to understand changes over time.

Some experts claim that physicians remain solely in charge of the key health care decisions, including prescriptions for medications, hospital admissions and discharges, and recommendations for surgery.[1] Other experts argue that physicians have lost their dominant position.[2–5] Still others say that physicians are in danger of becoming alienated workers.[6,7] Although many experts do believe that physician authority has declined in the past 40 years,[5,8–10] strong empirical evidence has not been presented.

Methods

Data from the American Association of Medical Colleges for both 1969 to 1970 and 1989 to 1990 were compared to determine if the representation of medical doctors within US medical schools has changed over time.[11,12] In 1969 to 1970, there were 117 schools listed in the Association's directory, compared with 144 in 1989 to 1990. A sample of 62 medical schools in the latter period was randomly selected, and data were collected on the proportion of personnel with medical or other degrees. Identical data were collected on the subset of 45 schools that were listed in the 1969–1970 directory.

Another measure of physician authority can be taken in various health associations. The highest degrees (MD or other) held by presidents of the American Public Health Association (APHA), the American Hospital Association, the American Physical Therapy Association, and the American Nursing Association were examined over time to determine changes in physician representation.[13–15] Data on the number of hospital physician managers and chief executive officers for selected years since 1952 were also considered. Because data were found in two separate sources and the variables were defined differently, comparisons had to be made carefully.[16,17]

The only measure of authority positions held by physicians in the wider society involves physician members of the US Congress.[18,19] There are 535 members of the House and Senate; thus, over any 10-year period, the total member-years served is 5350.

Results and Discussion

Table 1 shows that, overall in 1970, 65.6% of the sample of "key personnel" listed in the American Association of Medical Colleges directory held medical doctorates versus 72.8% in 1990. This evidence suggests that physicians have gained positions of authority in medical schools, at least over the past 20 years.

The American Medical Association has always been headed by a physician. In contrast, the American Physical Therapy Association and the American Nursing Association have never been headed by physicians. None of the 20 presidents of the American Hospital Association from 1970 to 1989 was a medical doctor, in contrast to 1930 to 1969, when 22 of 40 (55%) were physicians (Table 2). With

The author is with the Department of Rehabilitation Medicine, Tufts University School of Medicine, Boston, Mass.

Requests for reprints should be sent to J. Scott Osberg, PhD, 481 Beacon St, #44, Boston, MA 02115.

TABLE 1—Percentage of Medical School Positions Held by Medical Doctors, 1970 and 1990

	1970	1990
Administrative staff	48.9	50.2
Basic sciences faculty*	40.7	33.2
Clinical sciences*	87.5	92.5
Overall**	65.6	72.8

Note. Based on a sample of 45 medical schools selected from the *Association of American Medical Colleges Directory*, 1969–1970 and 1989–1990.[11,12]
*$P < .05$.
**$P < .001$.

TABLE 2—Presidents of the American Public Health Association (APHA) and the American Hospital Association (AHA) and Physician Years in the US Congress, by Decade

	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s
Percentage of physician presidents of AHA	---	---	---	70	50	40	60	0	0
Percentage of physician presidents of APHA	---	---	---	80	100	70	70	60	30
Physician-years in the US Congress	71	59	88	82	74	51	39	47	50

Note. The highest degree (MD or other) of presidents of AHA and APHA were collected over time (1 = MD, 0 = other). Data were taken from the *American Hospital Association Guide to the Health Care Field, 1987 Edition*,[13] Appendix D4, for 1929 to 1986; from recent years in *Hospitals*[14]; and from issues of *American Journal of Public Health*.[15] Data for physician-years in Congress from 1900 to 1959 were found in Burrow[18]; data for 1960 to 1989 were found in *Vital Statistics on Congress*.[19]

respect to the APHA, the change in physician domination has been less drastic. During the period from 1930 to 1969, 32 of 40 APHA presidents were medical doctors (80%), compared with only 9 of 20 (45%) from 1970 to 1989.

Physicians no longer dominate *the presidency* of APHA, yet other ungraphed results show that they still have a strong presence in the organization. In 1987, 50% of the 16-member executive board held medical degrees, and except for the president, who had a law degree, all six APHA officers held medical doctorates.

Although a physician has never headed the American Nursing Association, nursing is one area over which physicians have lost much control. Where nurses were once considered the physicians' "handmaidens," nursing now is a separate field with its own clearly demarcated lines of authority.[20]

The American Physical Therapy Association also has never had a physician as its president (personal communication with association staff, December 1991). However, the association was subservient to medicine, as seen in its mission statement in the 1937 Constitution: "To cooperate with, and to work only under the prescription of members of the medical profession."[21(p2)] In its early years, when it was called the American Women's Physical Therapy Association, it did have an advisory council comprised entirely of medical doctors.[22] However, this council was disbanded in 1968, and physicians no longer have a formal role in the association. In addition, in 1973 the association voted to bypass the need for physician referral.[23] The cumulative effect of all these changes indicates a loss of physician authority over physical therapy.[24]

There was a 13.5% decline in physician administrators, from 1851 in 1952 to 1602 in 1962.[17] According to Katzive, the heyday of medical administrators was in the early 1950s, with physician administrators representing 32.5% of all hospital administrators.[17] In contrast, 25.0% of administrators were physicians in 1945, and by 1962, although the number of hospital administrators grew, the percentage of physician administrators dropped to 22.8%. During the 10-year span beginning in 1972, the number of chief executive officers who were physicians declined by 75%, from 813 to 202.[16]

The trend away from physician administrators and hospital chief executive officers with medical degrees supports the large body of literature that has noted the rise of the hospital administrator and the decline of the physician administrator.[3,25–27] But is the flip side of administrators' rise the physicians' decline? As Freidson points out, even though nonphysician administrators hold many of the top posts, they are in many ways still beholden to the medical staff.[28] It could be that the rise of the hospital administrator is more properly considered a sign of a larger trend toward specialization,[29,30] and not a sign of an onslaught against physicians or of physician decline.

Table 2 shows physician membership in the US Congress in this century. In the 1920s, at the peak of physician membership, physicians constituted only 1.6% of the membership (88 of 5350 years). At the lowest point, in the 1960s, congressional membership comprised 0.7% physicians. Physicians may be less involved as national office holders, but perhaps Altman is right in noting that physicians today are more involved at the state and local level than they were in the past.[31]

Although today physicians make up a smaller proportion of Congress, they still have a great deal of influence over that body; US politicians have long been reluctant to offend the medical profession by seriously challenging physician authority.[32] In the first 18 months of the 1990 election cycle, the American Medical Association gave $2.1 million to congressional candidates.[33] It remains to be seen how physicians will use their political clout to shape President Clinton's proposed Health Security Act. Will Congress be able to resist the lobbying of physicians, hospitals, insurance companies, and pharmaceutical companies and pass meaningful reform?

These data reveal some losses in authority positions held by physicians over the past several decades; yet in medical schools, a remaining bastion of medical power, physicians seem to have solidified and expanded their hegemony. Further analysis of similar longitudinal data is needed, along with other research showing the extent to which physicians are pulling the strings behind the scenes from powerful, but less conspicuous, positions. □

References

1. Redisch MA. Physician involvement in hospital decision making. In: McKinlay JB, ed. *Health Care Consumers, Professionals, and Organizations*. Cambridge, Mass: MIT Press; 1988:337–363.
2. Ermann MD. The social control of organizations in the health care area. *Milbank Mem Fund Q*. Spring 1976;?:191–207.
3. Fuchs V. *Who Shall Live? Health Economics and Social Choice*. New York, NY: Basic Books Inc; 1983.

4. Starr P. *The Social Transformation of American Medicine.* New York, NY: Basic Books Inc; 1982.
5. Winkenwerder W, Ball JR. Transformation of American health care: the role of the medical profession. *N Engl J Med.* 1988;318:317–319.
6. Oppenheimer M. The proletarianization of the professional. *Sociol Rev [Monogr].* 1973;20:213–227.
7. McKinlay JB, ed. *Economic Aspects of Health Care; A Selection of Articles from the* Milbank Mem Fund Q. New York, NY: Prodist; 1973.
8. Krause EA. *Power and Illness: The Political Sociology of Health and Medical Care.* New York, NY: Elsevier; 1977.
9. Wolfe S. Professional dominance and the medical profession. *Am J Public Health.* 1986;76:11–12.
10. Reed RR, Evans D. The deprofessionalization of medicine: causes, effects, and responses. *JAMA.* 1987;258:3279–3282.
11. *AAMC Directory of American Medical Education, 1969–1970 Edition.* Washington, DC: Association of American Medical Colleges; 1970.
12. *AAMC Directory of American Medical Education, 1989–1990 Edition.* Washington, DC: American Association of Medical Colleges; 1990.
13. *American Hospital Association Guide to the Health Care Field, 1987 Edition.* Chicago, Ill: American Hospital Association; 1987.
14. *Hospitals.* Chicago, Ill: American Hospital Publishing, Inc; various years.
15. *Am J Public Health.* Washington, DC: American Public Health Association, Inc; various years.
16. Neuhauser D. *Coming of Age: A 50-Year History of the American College of Hospital Administrators and the Profession It Serves.* Chicago, Ill: Pluribus Press; 1983.
17. Katzive JA. The vanishing hospital administrator. *Hosp Topics.* 1965;49:41–42.
18. Burrow JG. *AMA: Voice of American Medicine.* Baltimore, Md: The Johns Hopkins University Press; 1963.
19. *Vital Statistics on Congress, 1990–91 Edition.* Washington, DC: American Enterprise Institute for Public Policy Research; 1991.
20. Eubanks P. Quality improvement key to changing nurse-MD relations. *Hospitals.* April 20, 1991;26–30.
21. APTA Task Force on Organizational Structure. History of the APTA structure. Presented to the board of the American Physical Therapy Association; November 1986. Exhibit 40-A, appendix C:1–17.
22. American Women's Physical Therapy Association [now the American Physical Therapy Association]. *The P.T. Review.* 1921.
23. Singleton MC. Independent practice: on the horns of a dilemma: a special communication. *Phys Ther.* 1987;67:54–57.
24. Navarro V. Professional dominance or proletarianization?: neither. *Milbank Mem Fund Q.* 1988;66(suppl 2):57–75.
25. Stevens R. *In Sickness and in Wealth: American Hospitals in the Twentieth Century.* New York, NY: Basic Books Inc; 1989.
26. McKinlay JB, Stoeckle JD. Corporatization and the social transformation of doctoring. *Int J Health Serv.* 1988;18:191–205.
27. Colombotos J, Kirchner C. *Physicians and Social Change.* New York, NY: Oxford University Press; 1986.
28. Freidson E. *Medical Work in America: Essays on Health Care.* New Haven, Conn: Yale University Press; 1989.
29. Snoke AW. *Hospitals, Health, and People.* New Haven, Conn: Yale University Press; 1987.
30. Hughes EC. *Men and Their Work.* Glencoe, Ill: The Free Press; 1958.
31. Altman LK. A view from Vermont where the governor is also the doctor. *New York Times.* September 3, 1991:C3.
32. Rovner J. Doctor bills are next target for cost-control efforts. *Cong Q Weekly Rep.* 1989:386–392.
33. Brightbill T. Political action committees: how much influence will $7.7 million buy? *Int J Health Serv.* 1991;21:285–290.

Estimates of Population Smoking Prevalence: Self- vs Proxy Reports of Smoking Status

Elizabeth A. Gilpin, MS, John P. Pierce, PhD, Shirley W. Cavin, MS, Charles C. Berry, PhD, Nicola J. Evans, PhD, Michael Johnson, PhD, and Dileep G. Bal, MD

ABSTRACT

Objectives. In the face of rising costs of surveillance systems, it is time to reexamine the feasibility of including proxy respondents in surveys designed to provide population estimates of smoking prevalence.

Methods. Data are from the California Tobacco Surveys, which are random-digit dialed telephone surveys. One adult provided demographic information and smoking status for all household residents. Additionally, some adults were selected for in-depth interviews that also included smoking status questions. We matched information from proxy respondents and self-respondents and evaluated smoking status discrepancies between them relative to demographic and other factors (n = 2930 matched pairs) in 1992. We address the potential bias these discrepancies might introduce into the population estimate of smoking prevalence.

Results. Overall, the discrepancy between proxy report and self-report was 4.3%, and it increased particularly when the self-respondent reported nondaily smoking or recent quitting. Discrepancies acted in both directions, and the net bias was that the screener survey overestimated smoking prevalence by 0.1% in 1992 (0.3% in 1990).

Conclusions. Smoking status questions can be added to ongoing surveys such as the census or labor force surveys; one adult could provide smoking status for all household members. (*Am J Public Health.* 1994; 84:1576–1579)

Introduction

Cigarette smoking has been identified as one of the most preventable causes of death and disability.[1,2] Accordingly, a surveillance system for smoking prevalence is important for many public health authorities at the state level. Establishing such a surveillance system requires addressing the issues of validity and cost.

The primary survey used to track smoking prevalence in the United States has been the National Health Interview Survey (NHIS).[2–4] Because the NHIS design provides population estimates representative of four regions of the continental United States, it cannot provide reliable state-specific estimates of smoking prevalence. In 1974 the NHIS decided to collect data on smoking status only from self-respondents.[5] Given that other types of surveys use proxy respondents, it is timely to consider the impact of including proxy respondents in the overall estimate of smoking prevalence.

Cost can be minimized if the proposed surveillance system can be incorporated into ongoing surveys, such as the population census or labor force participation surveys. In the United States, the Bureau of the Census conducts the Current Population Survey, which serves as the vehicle for evaluating state-level interventions in the large-scale American Stop Smoking Intervention Study project (ASSIST), which is a joint effort of the National Cancer Institute and the American Cancer Society.[6] These surveys typically seek information on all household members from one respondent; that is, they include information from a proxy respondent. Most households in the United States include two or more adults, and the second adult usually represents either a different sex or a different age group. In surveys that have sought to interview these other adults, the drop in response rate was considerable.[7] Thus, surveys that include information from proxy respondents may lead to more representative population estimates achieved with minimal statistical adjustment. Also, the sample size is increased considerably for only a marginal increase in the cost of a single interview. This advantage becomes particularly important when the survey is to be used to estimate smoking prevalence in demographic subgroups.

In the California Tobacco Surveys, one adult reported on his or her own smoking status and acted as a proxy respondent for other adults in the household. In addition, a subsequent in-depth survey of tobacco use was undertaken in some adults who had their smoking status previously reported by proxy on the screener survey.

This study examines the extent of discrepancies in current smoking status between the proxy–self pairs and explores demographic and other factors that might be associated with higher discrepancy levels. In addition, we computed adjusted estimates of smoking prevalence based on the observed discrepancy levels so that the

Elizabeth A. Gilpin, John P. Pierce, Shirley W. Cavin, and Nicola J. Evans are with the Cancer Prevention and Control Program, Cancer Center, and Charles C. Berry is with the Department of Family and Preventive Medicine, University of California–San Diego, La Jolla, Calif. Michael Johnson and Dileep G. Bal are with the Tobacco Control Section, California Department of Health Services, Sacramento, Calif.

Requests for reprints should be sent to John P. Pierce, PhD, Cancer Prevention and Control Program, Cancer Center, University of California–San Diego, La Jolla, CA 92093-0901.

This paper was accepted February 23, 1994.

degree of bias attributable to the inclusion of proxy responses could be determined.

Methods

Surveys

The California Tobacco Surveys are population-based, random-digit dialed telephone surveys.[8,9] In each household (n = 10 774 in 1992), an adult was administered a screener questionnaire that collected basic demographic information, including smoking status, for each household member. Each person enumerated (n = 21 870) was given a weight so that population-based estimates could be computed.[9] Briefly, the base weight reflects the probability that the household was selected, and it is adjusted to ensure that the sample is representative of the California population with respect to age, sex, county/region, education, and race/ethnicity.

From the information enumerated on the screener interview, some adults (≥ 18 years) were selected for in-depth interviews. All persons who had smoked in the last 5 years were selected, as were a random sample of 28% of those who had not (total n = 7263). A weighting procedure similar to that used for the screener respondents was also applied to the respondents who had the in-depth interview; it further adjusted for the fact that proportionately more smokers or persons who had quit smoking in the last 5 years than nonsmokers were interviewed in depth.[9] With this survey design, some individuals had smoking status reported both by proxy from the screener respondent and by self from the in-depth interview (n = 2930). It should be noted that 51.3% of the persons enumerated on the screener interview provided their own smoking status. In the analysis that examines the factors related to discrepancies (described below), the respondent weights from the in-depth interview are used.

Smoking Status

In the screener interview, the respondent was asked, "[As far as you know], have you [has person] smoked at least 100 cigarettes during your [his/her] lifetime?" If the answer to this question was "yes," the respondent was asked, "Do you [does person] smoke cigarettes now?" If the answer to this questions was "no," the respondent was asked, "Have you [has person] quit smoking during the past 5 years, [as far as you know]?"

In the in-depth interview, each respondent was asked the same set of questions. Those who had smoked 100 cigarettes and who were not smoking when surveyed were asked the date when they last smoked. In addition, current smokers were asked, "On how many days in the last month did you smoke?" Those smoking 25 or fewer days in the last month were considered occasional rather than daily smokers.[10] All persons who had smoked in the last year were asked the date and length of any quit attempts during the last year.

TABLE 1—Proxy-Reported Smoking Status for Each Group of Self-Reported Smoking Status

Smoking Status: Proxy Report	Smoking Status: Self-Report, No. (%)				
	Current Daily	Current Occasional	Quit Less than 1 Year Ago	Other[a]	Total
Current	911 (96.3)	154 (65.3)	46 (26.4)	59 (1.6)	1170
Quit less than 5 years ago	13 (1.2)	29 (10.5)	100 (49.6)	388 (19.2)	530
Other[a]	8 (2.5)	20 (24.2)	14 (24.0)	1188 (79.2)	1230
Total	932	203	160	1635	2930

Note. Percentages are weighted with in-depth interview weights (see Methods section).
[a]Other includes never smokers and longer-term former smokers.

Analytical Methods

Proxy–self pair discrepancies. For each category of self-reported smoking status (current daily, current occasional, quit in last year, nonsmoker for at least 1 year), we computed the weighted percentages of proxies reporting current smoking, quit in last 5 years, and nonsmoker for at least 5 years. A logistic regression of the likelihood of discrepancy was conducted to determine whether factors that appeared related to an increased discrepancy rate were independently related after controlling for all the other factors.[11] Because of the complex survey design, a jackknife resampling procedure was used to obtain variance estimates for the regression coefficients so that the independent significance of each variable could be assessed.[9]

Effect of proxy errors on estimate of smoking prevalence. To calculate the impact of discrepancies on the overall estimate of smoking prevalence, we estimated discrepancy percentages for each smoking status group that could be determined from the screener data: proxy-reported current smokers who self-reported nonsmoking (D1), proxy-reported quitters in the last 5 years who self-reported current smoking (D2), proxy-reported former smokers for at least 5 years who self-reported current smoking (D3), and proxy-reported never smokers who self-reported current smoking (D4). In order to obtain estimates that are as accurate as possible, data from the 1990 California Tobacco Surveys were pooled with the 1992 data to estimate these discrepancy rates. (The 1990 California Tobacco Surveys surveyed 32 125 households with 65 139 adults enumerated. A total of 24 296 adults received the extended interview, yielding 9384 self–proxy pairs. The 1990 survey did not ascertain the relationship of the self- and proxy respondents, so this paper reports mainly the 1992 results.) Screener weights for the discrepant pairs were summed across surveys and divided by the sum of the sums of weights for the two surveys.

The estimate of smoking prevalence from the screener survey was computed as follows:

$$p = \sum_{i=1}^{K} w_i \cdot s_i / \sum_{i=1}^{K} w_i$$

where $s_i = 1$ if the respondent is a current smoker and 0 otherwise, and w_i is the weight for a person enumerated on the screener. K is the total number of adults from the screener interview. The four discrepancy rates estimated from the proxy–self pairs were used to correct the prevalence estimate from the screener survey as follows:

$$p = \left[\sum_{i=1}^{n0} w_i \cdot s_i + (1 - \mathrm{D1}) \cdot \sum_{i=1}^{n1} w_i + \mathrm{D2} \cdot \sum_{i=1}^{n2} w_i + \mathrm{D3} \cdot \sum_{i=1}^{n3} w_i + \mathrm{D4} \cdot \sum_{i=1}^{n4} w_i \right] / \sum_{i=1}^{K} w_i$$

where $n0$ = number of self-respondents on screener (reported own smoking status and served as proxy respondent for other adult household members); $n1$ = number

TABLE 2—Proxy–Self Discrepancies in Subgroups Defined by Various Factors Related to the Household and Proxy and Self-Respondents

	No. Pairs	Total Percentage (n = 175) Discrepant[a]	P[b]
Overall	2930	4.3	
Age of self-respondent, y			
18–24	503	8.1	Ref
25–44	1411	3.8	.348
45+	1016	3.5	.245
Sex of self-respondent			
Male	1738	5.2	Ref
Female	1192	3.2	.451
Education of self-respondent, y			
<12	441	8.5	Ref
12+	2489	3.9	.327
Race/ethnicity of self-respondent			
White	2003	3.3	Ref
African American	138	2.7	.320
Hispanic	596	7.3	.562
Asian/other	193	2.3	.662
Relationship of self-respondent to proxy respondent			
Child	357	6.3	Ref
Parent/guardian	303	2.7	.671
Spouse/partner	1678	2.9	.170
Sibling	147	3.7	.443
Other relative	112	6.3	.295
Unrelated	322	8.5	.439
No. adults in household			
2	1711	2.4	Ref
3–4	1052	6.3	.547
5+	165	7.5	.303
Elapsed time between interviews			
≤1 wk	1626	4.8	Ref
>1 wk	1304	3.8	.263
Smoking status of proxy			
Nonsmoker	2213	4.0	Ref
Smoker	717	5.8	.429
Smoking status of self-respondent			
Never, quit for >12 mo	1645	1.4	Ref
Quit for 1–12 mo	138	22.7	.006
Quit in last month	22	45.8	.060
Occasional	203	35.0	.001
Daily	922	4.3	.267

[a]Percentages are weighted with in-depth interview weights (see Methods section).
[b]P value gives the independent significance of the variable, adjusting for all others. Variables were coded so that variable categories can be compared to a reference ("Ref").

of proxy-designated current smokers; $n2$ = number of proxy-designated former smokers within the last 5 years; $n3$ = number of proxy-designated former smokers who quit more than 5 years ago; and $n4$ = number of proxy-designated never smokers. Note that $K = n0 + n1 + n2 + n3 + n4$.

Results

Smoking Status Agreement Among Proxy–Self Pairs

Table 1 shows the number of proxy–self pairs and how they are distributed with respect to agreement and disagreement. The percentages shown are weighted column percentages of the distribution of proxy classification for each category of self-classification. Self-respondent and proxy-respondent reports of current daily smoking agreed in 96.3% of the cases. Percentages for the "Other" possible smoking status classification are also shown. Of particular interest is that proxy respondents classified as current smokers 26.4% of self-reported quitters in the last year.

Factors Related to Proxy–Self Discrepancy

Overall, among the 2930 proxy–self pairs for 1992, there were 175 (4.3%, weighted percentage) discrepancies concerning current smoker or nonsmoker status (Table 2). In 105 of these cases the proxy called a self-reported nonsmoker a current smoker, and in 70 cases the proxy called a self-reported smoker a nonsmoker. When weighted, these two discrepancies represent nearly equal percentages of the proxy–self pairs, 2.1% and 2.2%, respectively. In addition to the overall discrepancy rate for each demographic category, Table 2 shows a P value that was obtained by using specially coded variables contrasting each category to the indicated reference in the logistic regression analysis. This P value indicates the independent significance of a given category after adjusting for all the other variables.

Although there were trends for more discrepancies for younger, male, Hispanic, and less-educated self-respondents, none of these factors was significantly related to the likelihood of discrepancy. Other nonsignificant trends included higher discrepancy rates when the self-respondent and proxy-respondent were unrelated or when the relationship was other than spouse/partner, parent/guardian, or sibling. Also, discrepancy rates increased with larger household size.

The discrepancy rate was not related to the smoking status of the proxy nor to the time elapsed between the screener and in-depth interviews. Some persons selected for an in-depth interview were not available when the screener interview was administered and were contacted later. If over a week elapsed between interviews, the self-respondent's smoking status might change.

When the self-respondent reported quitting in the last year, the discrepancy rate was significantly greater than that in the reference group of those who had not smoked in the last year. Finally, self-respondents who were occasional smokers showed a significantly higher rate of proxy discrepancy.

Effect of Potential Discrepancy on Estimate of California Smoking Prevalence

Smoking prevalence for California estimated from the screener survey was 20.0 ± 0.7% in 1992. The error rates D1, D2, D3, and D4 were estimated to be 12.0%, 7.8%, 1.9%, and 2.4%, respec-

tively. The corrected smoking prevalence estimate with these percentages was 19.9%. Thus, the screener tended to overestimate smoking prevalence by 0.1% in 1992 (0.3% in 1990).

Discussion

The use of proxy respondents in population surveys of smoking status can substantially decrease the cost of obtaining a smoking prevalence estimate with minimal bias caused by potential error on the part of the proxy respondent. Inclusion of proxy respondents led to a more conservative estimate of smoking prevalence compared with the corrected estimates. The survey results are likely to be more representative because the problem of nonresponse of individuals selected for subsequent self-report does not occur. Furthermore, such a cluster sampling approach has been demonstrated to produce unbiased population estimates.[12]

The two types of discrepancies with respect to current smoking status tended to cancel one another; however, the net effect was a slight increase in the estimate of smoking prevalence when proxy respondents were included. As in a previous study,[13] the degree of discrepancy was not great, and it should be remembered that about half of the screener adults were self-respondents. The overall nature of the discrepancies was very similar for both the 1990 and 1992 surveys (Table 1), so if the same type of screener survey instrument is used for surveillance of smoking prevalence, any bias that is introduced should be present to the same degree. Thus, any change in smoking prevalence detected with such an instrument should be unaffected by the bias.

Although the proxy–self discrepancy rate was particularly high in some cases (nondaily smokers and recent quitters), these groups comprise a relatively small fraction of the total adult population. In agreement with a previous study,[13] we found that spouses and children tended to provide more accurate information than other relatives or unrelated individuals. This was expected as the immediate family would typically have a closer relationship with the individual and hence have more complete knowledge of his or her behavior. Hispanics appear to have a higher discrepancy rate than other racial/ethnic groups. Hispanic households are typically large, so household size may have partly accounted for this result. The more likely factor is occasional smoking, which we have previously shown to be particularly high among Hispanic smokers.[7,10]

We assumed that the self-report of smoking status was the correct one; however, it has been demonstrated that around 30% of self-reported quitters who participated in cessation programs have positive biochemical tests indicating that they have recently smoked.[14–16] Some programs validate self-report of recent quitting with the report of a significant other.[17] In our study, 26.4% of self-reported recent quitters were reported by the proxy to be smoking (Table 1). These reporting discrepancies are similar to those identified in clinic programs. Smokers are faced with considerable social pressure to quit, which may lead to a "wish bias" in self-reporting or desire to please the interviewer. Finally, those who are trying to quit or who do not smoke daily simply might not consider themselves smokers although the proxy does.

Although smoking prevalence is usually the most important issue that needs to be addressed in assessing a population's smoking problem, information on initiation, quitting, and relapse patterns is also of interest. Others have shown that proxy respondents can be reliable for current status of health issues,[18] and the screener survey instrument we described appears reliable for estimating smoking prevalence. However, because of concerns about accuracy,[19–23] we did not obtain detailed smoking histories and consumption levels from proxy respondents.

Our results from two large population surveys indicate that the use of proxy reporting of smoking status is cost-effective and minimally biased. Thus, it is recommended that surveillance systems for smoking prevalence use the approach of having one adult report smoking status for all adults in the household. □

Acknowledgment

This study was supported by contract 89-97872 from the California Department of Health Services, Tobacco Control Section, Sacramento, Calif.

References

1. *Smoking and Health. A Report of the Surgeon General.* Washington, DC: US Dept of Health, Education and Welfare; 1979. DHEW publication PHS 79-50066.
2. *Reducing the Health Consequences of Smoking: 25 Years of Progress. A Report of the Surgeon General.* Washington, DC: US Dept of Health and Human Services; 1989. DHHS publication CDC 89-8411.
3. *The Health Benefits of Smoking Cessation: A Report of the Surgeon General.* Washington, DC: US Dept of Health and Human Services; 1990:25–59. DHHS publication CDC 90-8416.
4. Pierce JP, Fiore MC, Novotny TE, Hatziandreu EJ, Davis RM. Trends in cigarette smoking in the United States: projections to the year 2000. *JAMA.* 1989;261:49–55.
5. National Center for Health Statistics. The National Health Interview Survey design, 1973–84, and procedures, 1975–83. *Vital Health Stat [1].* 1985;18. DHHS publication PHS 85-1320.
6. American Stop Smoking Intervention Study (ASSIST) for Cancer Prevention. December 1989. Proposal in response to RFP NCI-CN-95165-38.
7. Pierce JP, Farkas A, Evans N, et al. *Tobacco Use in California 1992. A Focus on Preventing Uptake in Adolescents.* Sacramento, Calif: California Department of Health Services; 1993.
8. Waksberg J. Sampling methods for random-digit dialing. *J Am Stat Assoc.* 1978;73:40–46.
9. Pierce J, Goodman J, Gilpin E, Berry C. *Technical Report on Analytic Methods and Approaches Used in the Tobacco Use in California, 1990–1991 Report.* Sacramento, Calif: California Department of Health Services; 1992.
10. Evans N, Gilpin E, Pierce JP, et al. Occasional smoking among adults: evidence from the California Tobacco Survey. *Tobacco Control.* 1992;1:169–175.
11. SAS Institute Inc. *SAS/STAT User's Guide,* Release 6.03 Edition. Cary, NC: SAS Institute Inc; 1988.
12. Cochran WG. *Sampling Techniques.* 2nd ed. New York, NY: John Wiley & Sons; 1963.
13. McLaughlin JK, Diety MS, Mehl EX, Blot WJ. Reliability of surrogate information on cigarette smoking by type of informant. *Am J Epidemiol.* 1987;126:144–146.
14. Ohlin P, Lundh B, Westling H. Carbon monoxide blood levels and reported cessation of smoking. *Psychopharmacology.* 1976; 49:263–265.
15. Brockway BS. Chemical validation of self-reported smoking rates. *Behav Ther.* 1978;9: 685–686.
16. Pierce JP, Dwyer T, DiGiusto E, et al. Cotinine validation of self-reported smoking in commercially run community surveys. *J Chronic Dis.* 1987;40:689–695.
17. Swartz JL. *Review and Evaluation of Smoking Cessation Methods: The United States and Canada, 1978–1985.* Washington, DC: Division of Cancer Prevention and Control, National Cancer Institute; 1987. NIH publication 87-2940.
18. Moore JC. Self/proxy response status and survey response quality: a review of the literature. *J Off Stat.* 1988;4:155–172.
19. Kolonel LN, Hirohata T, Nomura AMY. Adequacy of survey data collected from substitute respondents. *Am J Epidemiol.* 1977;106:476–489.
20. Rogot E, Reid D. The validity of data from next-of-kin in studies of mortality among immigrants. *Int J Epidemiol.* 1975;4:51–54.
21. Pickle LW, Morris Brown L, Blot WJ. Information available from surrogate respondents in case-control interview studies. *Am J Epidemiol.* 1983;118:99–108.
22. Lerchen ML, Samet JM. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol.* 1986;123:481–489.
23. Machlin SR, Kleinman JC, Madans JH. Validity of mortality analysis based on retrospective smoking information. *Stat Med.* 1989;8:997–1009.

Job Strain and Health-Related Quality of Life in a National Sample

Debra J. Lerner, PhD, Sol Levine, PhD, Sue Malspeis, SM, and Ralph B. D'Agostino, PhD

ABSTRACT

Objectives. Studies of the health effects of job strain have focused on morbidity and mortality as outcomes. This is the first study to examine the relationship of job strain to more comprehensive health status measures that encompass health-related quality of life.

Methods. In a national cross-sectional survey, 1319 working men and women, aged 18 through 64 years, completed a modified version of the Job Content Questionnaire that classified workers' jobs into four categories: high strain, passive, low strain, and active. Subjects also completed the Medical Outcomes Study Short-Form Health Survey and a health distress scale. Logistic regression analyses were employed that controlled for age, race/ethnicity, gender, and education. Both work and nonwork variables were included.

Results. Job strain was significantly associated with five of nine components of health-related quality of life: physical functioning, role functioning related to physical health, vitality, social functioning, and mental health. Job strain made a modest yet statistically significant contribution beyond the effects of chronic illness and psychosocial variables.

Conclusions. The results provide justification for further investigating the role of job strain as an independent risk factor for health-related quality of life. (*Am J Public Health.* 1994;84:1580–1585)

Introduction

A large literature on job strain[1] and its psychological and physical sequelae[1–11] has made us aware of the significant influence of the psychosocial work environment on health. According to Karasek and Theörell,[1] job strain and its health effects result from the interaction of two factors: (1) a job's psychological "demands" and (2) the degree of "control" or decision latitude a worker may exert over those demands. Control is, in effect, a modifier of a job's psychological demands.

In mostly cross-sectional and/or retrospective studies, "high-strain" jobs (high demand and low control) have been found to be negatively associated with health. Job strain has been associated with cardiovascular disease, left ventricular hypertrophy, and cardiovascular disease risk factors[2–7,11]; retiree mortality[10]; and sleep problems, depression, and psychosomatic symptoms.[8] Hypertension studies have produced conflicting results.[6,9] The "healthier" jobs have been characterized by low demand and high control (i.e., the "low strain" jobs) or high demand and high control (i.e., the "active" jobs). "Passive" jobs (low demand and low control) fall in between the two health extremes.

Our job-strain study goes beyond the usual morbidity or mortality outcomes by assessing the relationship of job strain to indicators of health-related quality of life. We also expand on health-related quality-of-life studies that focused on the influences of illness and medical care while excluding social variables.

Health-related quality-of-life indicators usually encompass a person's ability to function effectively physically, emotionally, and socially and to maintain a sense of well-being.[12–14] The more traditional biomedical health status measures capture the presence, absence, and/or severity of disease or mortality. Although health-related quality-of-life indicators were developed mainly to reflect the consequences of health problems and related treatments,[13] losses in functional capacity or sense of well-being may conceivably precede, follow, or be independent of disease.

We assessed job strain's contribution to health-related quality of life independently of demographic, medical, and other psychosocial variables; many previous studies did not consider other variables that might modify the job strain–health relationship. We also asked whether family and/or workplace supports buffer health differences attributable to job strain.

Methods

Our nationwide health survey was conducted between October 15 and December 22, 1990, by the National Opinion Research Center. Participants were drawn from the Research Center's own 1989 and 1990 General Social Survey samples. For our survey, the Research Center attempted to reinterview 3251 subjects from 2909 General Social Survey households. The resulting sample consisted of 2474 respondents (total response rate = 77%).

Debra J. Lerner, Sol Levine, and Sue Malspeis are with The Health Institute, New England Medical Center, Boston, Mass. Ralph B. D'Agostino is with the Department of Mathematics, Boston University, Boston, Mass.

Requests for reprints should be sent to Debra J. Lerner, PhD, The Health Institute, New England Medical Center, 750 Washington St, Box 345, Boston, MA 02111.

This paper was accepted January 25, 1994.

The present analysis was limited to 1319 men and women under 65 years of age who were employed full or part time. Most respondents at least 65 years of age were retired.

Health-Related Quality of Life

Health-related quality of life was measured with the 36-item version of the Medical Outcomes Study Short-Form Health Survey (SF-36).[15] The SF-36 consists of eight subscales measuring aspects of physical, emotional, social, and role functioning; bodily pain; vitality; and general and mental health perceptions. For most items, respondents are asked to evaluate their own health during the past 4 weeks. We also included a health distress scale developed by the same researchers.

The psychometric properties of the SF-36 have been examined extensively and support its validity, reliability, and precision and its advantages over similar instruments in detecting low-level health deficits.[16] It has been shown to distinguish healthy from chronically ill individuals, patients with different types of chronic conditions and degrees of severity, and medically and psychiatrically ill groups.[17–21] The internal-consistency reliability of most of the individual subscales based on the present data was excellent ($\alpha s \geq .76–.93$). However, the social function subscale, which contains only two items, did not perform as well ($\alpha = .62$).

SF-36 scores range from 0 (low) to 100 (high). A maximum score of 100 implies the absence of impairment. Results from the Medical Outcomes Study with the 20-item version of the SF-36 indicate that on the general health perceptions, mental health, or bodily pain subscales, a 3-point difference is sufficient to distinguish non-chronically ill persons from those with some form of chronic illness.[18]

In coding missing items on the SF-36 subscale, we followed the scoring method recommended by Ware and colleagues[22]: substituting the respondent's own mean subscale score for the missing item(s). For all other subscales, we used the more conservative approach of replacing a missing item score with the total sample's mean score for that specific item.

Measurement of the Independent Variables

A modified version of the Job Content Questionnaire[23] assessed job demands and control. (The item list can be obtained from the authors.) The Job Content Questionnaire was derived from the 1969, 1972, and 1977 National Quality of Employment Surveys, and recent analyses have begun to confirm its validity and reliability.[1] Cronbach's alpha statistics for the demand and control subscales used in this study were .71 and .76, respectively. One item in the control subscale, "repetitive work," had an item-to-total Pearson correlation coefficient of .21. ($\alpha = .81$). Principal-components analysis identified three factors. Two were roughly comparable to Karasek's original demand and control subscales. Repetitive work emerged as a third factor, although it contributed more heavily to the control factor. In keeping with Karasek's model, we elected to keep repetitiveness in the control subscale.

We created the four job-strain groups by dividing the distributions of the demand and control dimensions at their respective medians (thereby constructing a high and low group for each dimension) and cross-classifying subjects. Scores located on the demand and/or control medians were classified as "low." Figure 1 shows the resultant four job-strain groups.

We also included the following additional Job Content Questionnaire items:

- A physical demand item.
- Two items measuring support, one each from coworkers and supervisors.
- Single-item measures of perceived job security, skill obsolescence, and future ambiguity (the last three were summed to form an index of job insecurity).

The following additional classes of independent variables were included:

- Demographics.
- Employment status.
- Chronic illness.
- Psychosocial variables (e.g., social supports,[24] life satisfaction).

Descriptive data for these variables are included in Table 1.

Statistical Analysis

We produced descriptive data for the major variables (Table 1), statistical comparisons of health status scores among the job-strain groups (Table 2), and multivariate models depicting the relationship of each of the nine health subscales to the independent variables (Tables 3 and 4). Three dependent variables (vitality, general health perceptions, and mental health) were normally distributed. Consistent with previous studies, the remaining six dependent variables were nonnormal, with the bulk of the responses at 100.[18]

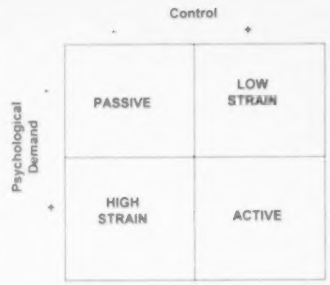


FIGURE 1—Job-strain model.

TABLE 1—Demographic and Psychosocial Characteristics of Study Sample (n = 1319)

Characteristic	Percentage
Female	51.9
White	83.5
Not married	39.7
Full-time job	83.9
Job type	
Manual	22.4
Nonmanual	43.4
Other/sales and service	34.2
No. chronic conditions[a]	
None	34.3
1	25.0
2	19.3
3 or more	21.4
	Mean
Age, y	39.0
Education, y	13.6
General social support (min = 5, max = 25)	18.9
Satisfaction, financial and environmental conditions (min = 4, max = 20)	13.6
Satisfaction, psychosocial conditions (min = 2, max = 10)	7.3
General life satisfaction (min = 1, max = 5)	3.4

Note. Sample includes full-time and part-time employees younger than 65 years of age. No more than 10 responses were missing from any variable.

[a]To create the chronic-condition variable, we combined responses from two questions: whether a doctor had diagnosed any of six specific conditions (yes = 21%) and a self-report of the presence or absence of eight specific conditions and an "other condition" (yes = 61%).

TABLE 2—Mean Health Status Scores, by Job-Strain Group

Health Measures	Job-Strain Groups[a]										
	Passive (n = 430)		Low Strain (n = 303)		High Strain (n = 314)		Active (n = 260)		Total Sample (n = 1307)		US Norm[b]
Physical function[c]	90.0	1.5	92.9	1.5	85.3	2.3	93.0	1.7	90.1	0.9	84.1
Role function/physical[c]	88.8	2.5	90.9	2.6	80.1	3.8	91.7	2.6	87.8	1.5	81.0
Bodily pain[c]	77.0	2.0	81.3	2.2	71.2	2.7	80.8	2.2	77.4	1.2	75.2
General health perceptions[d]	74.4	1.6	79.2	1.8	71.1	2.3	80.2	1.9	75.9	1.0	72.0
Vitality[d]	62.0	1.8	67.6	1.8	54.4	2.5	64.6	2.4	62.0	1.1	60.9
Social function[c]	86.5	1.9	87.6	2.3	79.4	2.6	88.7	2.2	85.5	1.1	83.3
Role function/emotional[c]	84.5	2.3	88.1	2.9	75.6	4.1	86.3	3.4	83.6	1.7	81.3
Mental health[d]	75.5	1.6	79.7	1.7	68.5	2.2	77.2	1.9	75.1	0.9	74.7
Health distress[c]	87.4	1.8	91.0	1.6	81.5	2.7	89.6	2.0	87.3	1.0	85.3

Note. Numbers in columns represent mean and 95% CI (±) with $P < .001$ for all row comparisons.
[a]No more than four responses were missing from any job-strain group.
[b]National norms based on the National Opinion Research Center survey are means weighted to reflect age, sex, and regional distribution of the US population.
[c]Differences tested with Kruskal-Wallis test.
[d]Differences tested with analysis of variance.

Health status differences among the job-strain groups were tested with analysis of variance (ANOVA) for the normally distributed dependent variables. The Kruskal–Wallis test was used for the non-normally distributed health measures (Table 2).

Before final logistic regression analyses were performed, a series of cross-tabulations, correlations, and logistic stepwise regressions were run to clarify the relationships of independent variables within the same conceptual class (e.g., demographic) to the dependent variables. This analysis helped minimize redundancy among conceptually related independent variables. Bivariate correlations (Pearson's Product Moment Correlation) showed total social support to be more strongly related to the dependent variables than either the affective-support or the instrumental-support variable. Marital status did not add significantly to the demographic model but was retained as an independent variable because of its substantive importance. The full set of nonredundant variables was entered into logistic regression models. We culled the statistically significant variables for our final models and identified any variables causing collinearity problems (see Results). Health scores were dichotomized at the median, and we controlled for age, gender, race/ethnicity, and education. Job strain was entered as a four-class dummy variable with the high-strain group always used as the comparison group and coded as zero.

Results are presented in Tables 3 (physical health dimensions) and 4 (emotional health dimensions) as odds ratios and associated significance levels, model chi squares and associated significance levels, and areas under the receiver operating characteristic (ROC) curve. ROC curves plot the relationship between the rate of true-positive classifications to the rate of false-positive classifications.[25] The area under the curve is a goodness-of-fit measure with values above .75 considered acceptable. All models were checked for collinearity problems and none were found.

Results

The demographic characteristics of the sample are described in Table 1.

Work Characteristics

Slightly less than one fourth (23%) of the workers in our sample had high-strain jobs. One third had passive jobs, 24% had low-strain jobs, and one fifth had active jobs.

Almost one fourth (24%) had physically demanding jobs. However, although many described work as psychologically and/or physically demanding, most coworkers (80%) and supervisors (65%) were considered supportive.

Sixteen percent reported job insecurity, and 28% regarded their career prospects as poor. Only 5% predicted their own skill obsolescence.

SF-36 Subscales (Table 2)

Mean SF-36 scores were consistently lowest among workers in the high-strain jobs. The active group and the low-strain group alternated as being the healthiest of the four. The passive group was in the middle.

Health disparities among the job-strain groups were sometimes large. There was a 13.2-point difference in mean vitality scores between the high-strain and low-strain groups and a 12.5-point difference in mean role function/emotional scores (i.e., due to emotional health). All row comparisons with ANOVA or the Kruskal-Wallis test were statistically significant.

Multivariate Results (Tables 3 and 4)

The passive, low-strain, and active job groups each had statistically significant health advantages over the high-strain groups. After the influences of demographics, chronic conditions, and other variables were taken into account, high job strain was significantly associated with lower scores on five subscales: physical function, role function/physical (i.e., due to physical health), vitality, social function, and mental health.

Compared with the high-strain group, the low-strain and passive groups had about 60% to 80% higher odds of experiencing better physical function, role function/physical, social function, and vitality. Vitality levels in the active group also were significantly better than those in the high-strain group. Job strain had its largest impact on mental health; compared with the passive and low-strain groups, the high-strain group had double the odds of poorer mental health.

TABLE 3—Logistic Regression Models: Odds Ratios (95% Confidence Intervals) for Physical Health Dimensions

Independent Variable[a]	Physical Function: High (95–100) vs Low (0–94)	Role Function/ Physical: High (100) vs Low (0–99)	Pain: High (84–100) vs Low (0–83)	General Health Perceptions: High (77–100) vs Low (0–76)	Vitality: High (65–100) vs Low (0–64)
Job strain (9)					
High strain	1.00	1.00			1.00
Passive	1.32 (0.92, 1.88)	1.76 (1.20, 2.58)			1.59 (1.13, 2.25)
Low strain	1.80 (1.19, 2.72) *	1.48 (0.95, 2.30) *			1.62 (1.10, 2.40) *
Active	1.35 (0.89, 2.06)	1.34 (0.84, 2.13)			1.64 (1.09, 2.45)
Workplace support (1)			1.08 (1.01, 1.17)		1.11 (1.03, 1.21)
Strain × workplace support					
Social support (2)	1.04 (1.01, 1.07)	1.06 (1.02, 1.09)	1.04 (1.01, 1.07)		
Strain × social support					
No. chronic conditions (3)	0.53 (0.47, 0.60)	0.48 (0.41, 0.55)	0.50 (0.44, 0.56)	0.60 (0.54, 0.68)	0.70 (0.62, 0.79)
Strain × no. chronic conditions					
Job insecurity (4)				0.91 (0.86, 0.96)	
Strain × job insecurity					
Physical job demands (5)	0.89 (0.79, 0.99)		0.78 (0.70, 0.86)	0.89 (0.81, 0.99)	
Life statisfaction (6)			1.26 (1.09, 1.45)		2.09 (1.77, 2.46)
Satisfaction/material (7)	1.05 (1.01, 1.10)	1.12 (1.07, 1.18)		1.07 (1.02, 1.12)	1.07 (1.02, 1.12)
Satisfaction/social (8)				1.21 (1.11, 1.31)	1.13 (1.03, 1.24)
Married (10)					0.65 (0.49, 0.86)
Manual work (11)					
Nonmanual work (12)					
Full time (13)					
No.	1275	1281	1284	1286	1275
Receiver operating characteristic area	.78	.77	.77	.75	.78
Model chi-square	323.42**	225.54**	304.17**	261.44**	333.29**
df	11	10	9	9	12

Note. Models control for continuous age; gender (male = 1, female = 0); race (White = 1, non-White = 0); and continuous education.

[a]Continuous variables and ranges: (1) workplace support (2 to 10); (2) social support (5 to 25); (3) number of chronic conditions (0 to 3 or more); (4) job insecurity (3 to 15); (5) physical job demands (1 to 5); (6) life satisfaction (1 to 5); (7) satisfaction/material (4 to 20); (8) satisfaction/social (2 to 10). Dichotomous variables: (9) job strain, entered as a set of three dummy variables with high strain as the comparison group; (10) 1 = married, 0 = not married; (11) 1 = manual work, 0 = professional, sales, clerical, service work; (12) 1 = nonmanual work, 0 = manual, sales, clerical, service work; (13) 1 = full time, 0 = part time.

*$.01 < P \leq .05$; **$P \leq .001$.

Social Supports (Tables 3 and 4)

The main effects of workplace support were associated with higher odds of freedom from bodily pain as well as improved vitality, and role function, and emotional health. General social support was associated with five of the subscales. However, none of the strain and support interaction terms, which would have suggested a buffering effect, was significant. In early analyses, the interaction term "workplace support by job strain" had a significant negative association with role function/emotional; the high-strain, low-support group had the lowest scores. However, the interaction term was collinear with the main effect of workplace support and was dropped from the model. Collinearity problems caused by the general "social support by job strain" interaction term in the physical-function model led to its elimination.

Other Variables

Two thirds of the sample reported one or more chronic conditions. The number of chronic conditions reported (up to three or more) was strongly associated with all nine subscales. Each unit increase in the number of chronic conditions decreased the odds of good health by one fourth to one half.

Job insecurity and physically demanding work were associated with poorer health in several models. Life satisfaction had a strong, positive association in five of the eight models (mental health was excluded). People who reported high satisfaction were twice as likely to have good vitality, social function, and role function/emotional.

Satisfaction with environmental and financial circumstances (e.g., housing, standard of living) and satisfaction with family and friends were linked significantly with health in several of the models.

Discussion

Job strain was associated with five components of health-related quality of life: physical functioning, role functioning related to physical health, vitality, social functioning, and mental health. Some health differences, such as those contained in the mental health model, were substantial. In contrast, job strain was associated only marginally with physical functioning ($P \geq .051$).

Although the results support our main hypotheses, we were limited by cross-sectional data. Also, both the SF-36 and job-strain variables were based on self-report. Nevertheless, the SF-36 is associated with a range of objectively and subjectively assessed health measures[13,17,20] and thus should be relatively free of bias.

TABLE 4—Logistic Regression Models: Odds Ratios (95% Confidence Intervals) for Emotional Health Dimensions

Independent Variable[a]	Social Function: High (100) vs Low (0–99)	Role Function/ Emotional: High (100) vs Low (0–99)	Mental Health: High (80–100) vs Low (0–79)	Health Distress: High (95–100) vs Low (0–94)
Job strain (9)				
High strain	1.00		1.00	
Passive	1.81 (1.29, 2.53)		2.25 (1.59, 3.19)	
Low strain	1.68 (1.15, 2.44)	*	2.08 (1.40, 3.08)	**
Active	1.32 (0.89, 1.94)		1.53 (1.02, 2.30)	
Workplace support (1)		1.15 (1.06, 1.25)		
Strain × workplace support				
Social support (2)	1.04 (1.01, 1.07)		1.13 (1.09, 1.16)	
Strain × social support				
No. chronic conditions (3)	0.72 (0.64, 0.81)	0.74 (0.65, 0.84)	0.71 (0.62, 0.80)	0.53 (0.47, 0.60)
Strain × no. chronic conditions				
Job insecurity (4)			0.91 (0.85, 0.97)	0.89 (0.84, 0.94)
Strain × job insecurity				
Physical job demands (5)	0.88 (0.80, 0.98)			0.88 (0.79, 0.98)
Life satisfaction	1.83 (1.58, 2.11)	1.94 (1.64, 2.30)		1.44 (1.25, 1.66)
Satisfaction/material (7)				1.08 (1.03, 1.13)
Satisfaction/social (8)		1.12 (1.02, 1.23)	1.38 (1.26, 1.50)	
Married (10)		1.37 (1.03, 1.84)	0.76 (0.58, 1.01)	
Manual work (11)				
Nonmanual work (12)				
Full time (13)				
No.	1282	1275	1277	1286
Receiver operating characteristic area	.75	.76	.78	.76
Model chi-square	255.39**	235.80**	350.07**	286.68**
df	11	8	11	9

Note. Models control for continuous age; gender (male = 1, female = 0); race (White = 1, non-White = 0); and continuous education.

[a]Continuous variables and ranges: (1) workplace support (2 to 10); social support (5 to 25); (3) number of chronic conditions (0 to 3 or more); (4) job insecurity (3 to 15); (5) physical job demands (1 to 5); (6) life satisfaction (1 to 5); (7) satisfaction/material (4 to 20); (8) satisfaction/social (2 to 10). Dichotomous variables: (9) job strain, entered as a set of three dummy variables with high strain as the comparison group; (10) 1 = married, 0 = not married; (11) 1 = manual work, 0 = professional, sales, clerical, service work; (12) 1 = nonmanual work, 0 = manual, sales, clerical, service work; (13) 1 = full time, 0 = part time.

*$.001 < P \leq .01$; **$P \leq .001$.

We were also hindered by the nonparametric distributions of six of the dependent variables, which necessitated a logistic analysis. Thus, we know that job strain is useful in differentiating the "very healthy" from everyone else in the sample. However, we are not able to make more refined comparisons.

We also add an important caveat. The relationships between life satisfaction (i.e., feeling good or bad about one's circumstances) and some of the subscales must be viewed critically because the independent variable may be conceptually overlapping with some dependent variables.

Despite the fact that job strain reached levels of statistical significance in five models, it was sometimes less significant than other variables. We speculate that results may reflect an indirect pathway linking job strain to health-related quality of life. Job strain may operate directly on health-related quality of life by influencing one's perception of functional ability and well-being. However, job strain may also operate on functional ability and sense of well-being indirectly by contributing to the production of physical or mental disease; that is, health-related quality of life is a response to disease or its precursors. Thus, job strain may operate as both a proximal variable, having a relatively direct and immediate influence on health-related quality of life, and as a distal variable operating through the more proximal or mediating influences of stress-induced disease. The latter statement is consistent with the job-strain hypotheses constructed by Karasek, Theorell, and others.[1]

The lack of a significant interaction between job strain and workplace support or general social support indicates that there were no significant buffering effects. The buffering role of workplace support remains controversial,[26] and recent studies have produced results conflicting with our own.[10] Our workplace-support variables may have been too narrowly defined to be statistically significant.

Establishment of job strain as a risk factor for health-related quality of life awaits further study. Studies of the health effects of chronic, long-term exposure to high-strain jobs, in accordance with recent research on Swedish workers by Johnson and Stewart,[27] suggest that it would be potentially fruitful to measure job strain at multiple points in time. In addition, we were unable to assess cumulative strains on worker health by including data pertaining to marital and parenting roles, financial problems, or other factors. However, analyses of the interplay among work, family, and community would strengthen research on both job strain and health-related quality of life.

Our study argues for continued attention to job roles as health risks. Because jobs are socially structured, the additional

risks associated with high-strain jobs should be considered preventable. □

Acknowledgments

Funding was provided by the New England Medical Center Inc's Society and Health Program, supported entirely by The Henry J. Kaiser Family Foundation.

The authors wish to acknowledge the following people for their helpful advice and support: Alvin R. Tarlov, MD, Colleen A. McHorney, PhD, John E. Ware Jr, PhD, Peter Schnall, MD, Paul Landsbergis, EdD, and Constance L. Kelley.

References

1. Karasek RA, Theorell TT. *Healthy Work: Stress, Productivity, and the Reconstruction of Working Life*. New York, NY: Basic Books; 1990.
2. Karasek RA, Baker D, Marxer F, Ahlbom A, Theorell TT. Job decision latitude, job demands and cardiovascular disease: a prospective study of Swedish men. *Am J Public Health*. 1981;71:694–705.
3. Alfredsson L, Karasek RA, Theorell TT. Myocardial infarction risk and psychosocial environment: an analysis of the male Swedish working force. *Soc Sci Med*. 1982;3: 463–467.
4. Theorell TT, Hamsten A, de Faire U, Orth-Gomer K, Perski A. Psychosocial work conditions before myocardial infarction in young men. *Int J Cardiol*. 1987;15:33–46.
5. Karasek RA, Theorell TT, Schwartz J, Schnall P, Pieper C, Michela JL. Job characteristics in relation to the prevalence of myocardial infarction in the US HES and HANES. *Am J Public Health*. 1988;78: 910–918.
6. Schnall PL, Pieper C, Schwartz JE, et al. The relationship between "job strain," workplace diastolic blood pressure, and left ventricular mass: index results of a case control study. *JAMA*. 1990;263:1929–1935.
7. Schnall PL, Schwartz JE, Landsbergis PA, et al. Relation between job strain, alcohol, and ambulatory blood pressure. *Hypertension*. 1992;19:488–495.
8. Landsbergis PA. Occupational stress among health care workers: a test of the job demands-control model. *J Organ Behav*. 1988;9:217–239.
9. Albright CL, Winkleby MA, Ragland DR, Fisher J, Syme SL. Job strain and prevalence of hypertension in a biracial population of urban bus drivers. *Am J Public Health*. 1992;82:984–989.
10. Falk A, Hanson BS, Isacsson S, Östergren P. Job strain and mortality in elderly men: social network, support, and influence as buffers. *Am J Public Health*. 1992;82:1136–1139.
11. LaCroix AZ. *Occupational Exposure to High Demand/Low Control Work and Coronary Heart Disease in the Framingham Cohort*. Chapel Hill, NC: University of North Carolina; 1984. PhD dissertation.
12. Levine S, Croog SH. What constitutes quality of life? A conceptualization of the dimensions of life quality in healthy populations and patients with cardiovascular disease. In: Wenger NK, Mattson ME, Furberg CD, Elinson J, eds. *Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies*. New York, NY: LeJacq Publications; 1984:46–66.
13. Stewart AL, Ware JE Jr. *Measuring Functioning and Well-Being—The Medical Outcomes Study Approach*. Durham, NC: Duke University Press; 1992:6.
14. Mosteller F, Ware JE Jr, Levine S. Advances in health status assessment: conference proceedings. *Med Care*. March 1989; 27(suppl):S282–S294.
15. Ware JE Jr, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36), I: conceptual framework and item selection. *Med Care*. 1992;30:473–483.
16. Brazier JE, Harper R, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 Health Survey questionnaire: new outcome measure for primary care. *BMJ*. 1992;305:160–164.
17. Stewart AL, Greenfield S, Hays S, et al. Functional status and well-being of patients with chronic conditions: results from the Medical Outcomes Study. *JAMA*. 1989; 262:907–913.
18. McHorney CA, Ware JE Jr, Rogers W, Raczek AE, Lu JFR. The validity and relative precision of MOS Short- and Long-Form Health Status Scales and Dartmouth COOP Charts: results from the Medical Outcomes Study. *Med Care*. May 1992;30(suppl):MS253–MS265.
19. Kravitz RL, Greenfield SL, Rogers W, et al. Differences in the mix of patients among medical specialties and systems of care: results from the Medical Outcomes Study. *JAMA*. 1992;267:1617–1623.
20. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36), II: psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993; 31:247–263.
21. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care*. 1992;30:917–925.
22. International Resource Center for Health Care Assessment. *How to Score the SF-36 Health Status Survey*. Boston, Mass: New England Medical Center; 1991.
23. Karasek RA. *The Job Content Questionnaire*. Los Angeles, Calif: University of Southern California, Department of Industrial and Systems Engineering; 1985.
24. Sherbourne CD, Stewart AL. The MOS Social Support Survey. *Soc Sci Med*. 1991; 32:705–714.
25. Pagano` M, Gauvreau K. *Principles of Biostatistics*. Belmont, Calif: Duxbury Press; 1993.
26. House JS. *Work Stress and Social Support*. Reading, Mass: Addison-Wesley Publishing Co; 1981.
27. Johnson JV, Stewart WF. Measuring work organization exposure over the life course with a job-exposure matrix. *Scand J Work Environ Health*. 1993;19:21–28.

Alcohol and Drug Disorders among Physically Abusive and Neglectful Parents in a Community-Based Sample

Kelly Kelleher, MD, MPH, Mark Chaffin, PhD, Janice Hollenberg, MS, and Ellen Fischer, PhD

ABSTRACT

Objectives. The purpose of this study was to compare the frequency of substance use disorders and symptoms between adults reporting child physical abuse or neglect and individually matched control subjects in a community sample.

Methods. In a nested case–control study, 169 adults reporting physical abuse of a child and 209 adults reporting neglect of a child from 11 662 individuals successfully interviewed in a probabilistic survey in four communities were individually matched with control subjects drawn from the participants. Case subjects were compared with control subjects on the number of alcohol- or drug-related symptoms and disorder diagnoses as determined by symptoms from the Diagnostic Interview Schedule.

Results. Respondents reporting either physical abuse or neglect of children were much more likely than their matched control subjects to report substance abuse or dependence. These differences persisted after potential confounding variables were controlled for.

Conclusions. Parental substance abuse and dependence, independent of confounding factors, are highly associated with child maltreatment. Inconsistent results in previous studies may have arisen from reliance on referred samples and unstandardized assessment methods. Agencies involved in the care of abused or neglected children and their families should consider incorporating routine substance abuse evaluations with treatment, or referral for treatment, where indicated. (*Am J Public Health.* 1994;84:1586–1590)

Introduction

Over the past three decades, the number of reports suggesting a relationship between child physical abuse or neglect and substance use by parents has grown.[1] Yet a recent national survey of state child protective services agencies revealed that only 14 states routinely collect information on substance abuse.[1] Estimates of the proportion of confirmed child abuse cases associated with some sort of parental substance abuse problem have varied widely, from 5% to 78%.[2] Nevertheless, substance abuse has been termed the "dominant characteristic" of families in the child abuse and neglect caseloads of some states.[3]

Social and environmental factors have been considered primary to the development of both child maltreatment and parental substance abuse. Social class is strongly linked to both reported child maltreatment[4] and a wide range of mental disorders, including drug and alcohol disorders.[5,6] Child physical abuse and neglect have been found to be higher by factors of 3.5 and 8.0, respectively, among lower-income families.[7]

Although social and environmental factors have been considered the primary determinants of child abuse and neglect, some authors[4,8,9] have suggested that parental psychiatric disorder generally, and substance abuse specifically, predisposes to child maltreatment, regardless of social factors. For example, incapacitation or impaired judgment secondary to substance abuse may give rise to child neglect. Similarly, depression has a documented association with child maltreatment.[9] Likewise, antisocial personality traits are associated with substance abuse and child maltreatment.[10] Finally, abusive parents are widely cited as being socially isolated,[11–13] and thus prone to drinking or drug use.

Methodological and sampling problems have left unanswered questions about the strength and nature of the relationship between alcohol and drug disorders and child maltreatment, thereby calling into question the value of increasing efforts targeting alcohol or drug abuse in families referred for child abuse or neglect. Studies comparing *reported* child abusers with their nonabusive counterparts are vulnerable to confounding as a result of the child abuse reporting process, which may disproportionately identify women, people of color, or lower socioeconomic classes. Studies examining rates of child abuse or neglect among clinical populations of substance-abusing parents are subject to similar referral biases.

Moreover, many studies used poorly specified or inconsistent definitions of substance abuse. Often, investigators did not distinguish among substance abuse (a general pattern of maladaptive use with associated social or physical problems), substance dependence (a more specific chronic pattern characterized by a wide

At the time of the study, all authors were with the University of Arkansas for Medical Sciences, Little Rock, Ark: Kelly Kelleher was with the Department of Pediatrics and the Department of Psychiatry and Behavioral Sciences; Mark Chaffin is with the Department of Pediatrics; Janice Hollenberg and Ellen Fischer are with the Department of Psychiatry and Behavioral Sciences. Ellen Fischer is also with the VA HSR&D Field Program for Mental Health, Little Rock.

Requests for reprints should be sent to Kelly Kelleher, MD, MPH, Children's Hospital of Pittsburgh, 3705 5th Ave, Room G205, Pittsburgh, PA 15213.

This paper was accepted February 4, 1994.

variety of role impairments, tolerance, withdrawal, and loss of control over use),[13-15] and a variety of lesser patterns of substance use. Some authors[12] suggested that reports of drinking found among surveys of abusive parents may simply reflect attempts to present a socially acceptable justification for abusive behavior, rather than reflecting the correspondence between child maltreatment and a bona fide diagnosis of addictive disease.

These concerns led some to conclude that the relationship between substance abuse and child maltreatment is minimal at best.[16] Yet we are aware of no studies that examined the epidemiology of alcohol and drug disorders with physical abuse and neglect that used (1) a probability sample from a community population rather than referred cases, (2) a carefully matched control group, and (3) data-analytic procedures able to adjust for potential confounds.

Methods

The National Institute of Mental Health (NIMH) Epidemiological Catchment Area Study was a community-based survey of 18 000 adults in five communities across the United States. The Epidemiological Catchment Area data provided the best estimates of the prevalence of other psychiatric disorders in the community, and they offer a unique opportunity to compare the frequency of substance abuse and dependence between adults who report having abused or neglected children and those not reporting child maltreatment. Adult respondents reporting abusive or neglectful behaviors (case subjects) were individually matched on sociodemographic factors with adult respondents not reporting those behaviors (control subjects). We hypothesized that abusive adults and neglectful adults would have more alcohol and drug disorders and symptoms related to substance abuse than their control subjects.

Sample and Procedures

The Epidemiological Catchment Area program consisted of sample surveys at five research sites: New Haven, Conn; Baltimore, Md; St. Louis, Mo; Durham, NC; and Los Angeles, Calif. The surveys were coordinated by the NIMH program staff and included a probability sample of 4000 noninstitutionalized individuals from each community.

Once an individual was selected for the sample, repeated attempts as necessary were made to interview the individual. Of the individuals sampled, 65% to 80% were successfully interviewed. Participation rates varied minimally by site, race, sex, and age.[17] The survey consisted of a face-to-face structured interview, using the Diagnostic Interview Schedule,[18] to obtain information on sociodemographics, health service use, psychiatric symptoms, and functioning.

Measures

Cases for this study were defined as all adult respondents who answered "yes" to any one of five questions on abusive or neglectful behaviors included in the Diagnostic Interview Schedule. More specifically, a single question in the Schedule inquires about serious physical abuse of a child resulting in bruises, bed days, or medical care. Participants who responded positively on this item were classified as "abusive." One site (New Haven)[13] did not include this question and was excluded from all analyses.

Four questions inquired about neglectful behaviors, including leaving very young children unattended for extended periods, inadequately feeding or caring for children, or having a health care professional state that children were being neglected. The respondents who reported at least one of these behaviors were classified as "neglectful."

All of the remaining respondents who reported never having had children or never having acted as caretakers for children were then excluded from the potential control group. The abusive adults and the neglectful adults were matched with respondents from the control pool on age (±5 years), race, gender, site, and socioeconomic status quartile. Socioeconomic status scores were determined from a composite of variables including education, household income, and occupational status. Eleven (3.0%) of the abusive or neglectful adults who could not be matched on race were eliminated from analysis. When more than one potential match was generated for a case, one match was randomly selected from among them.

The other primary variables of interest were alcohol and drug disorders. The Diagnostic Interview Schedule used in this study provided computer-generated diagnostic algorithms to derive lifetime *Diagnostic and Statistical Manual of Mental Disorders*, third edition, alcohol abuse, alcohol dependence, drug abuse, and drug dependence diagnoses. Subjects meeting these *Diagnostic and Statistical Manual of Mental Disorders* criteria were coded as positive for the relevant diagnoses. Previous work demonstrated the reliability and validity of the Schedule in detection of alcohol and drug disorders.[19] In addition, the number of alcohol disorder symptoms (out of a possible 17) and drug disorder symptoms (out of a possible 8) was calculated for each respondent. Because of the lack of information about the timing of child abuse or neglect behaviors, lifetime prevalence estimates instead of current prevalence estimates are reported for this study.

Because depression[20] and social isolation[21] have been associated with child abuse and neglect, lifetime history of major depressive disorder (yes/no) and social support, defined as use of friends and/or relatives as confidants (yes/no) or use of a social service agency for problems with emotions, nerves, drugs, alcohol, or other mental health problems (yes/no), were coded from the Diagnostic Interview Schedule. Because antisocial personality has been associated with excessive drinking and drug use,[22] the presence of antisocial personality was entered in regression models to examine the extent to which substance abuse disorders independently contribute to child physical abuse and neglect.

The number of individuals in a household was included in the regression model to control for the effect of the cluster sampling scheme, which oversampled individuals living alone, and to control for the demonstrated association between family size and child physical abuse and neglect.[7] Finally, to screen for a general "yes" response set, cases were assessed for the presence or absence of another disorder diagnosed by the Diagnostic Interview Schedule. Panic disorder was chosen because there was no theoretical or empirical evidence to suggest a relationship between panic disorder and child maltreatment.

Analytic Plan

Statistical methods for matched-pairs data were employed. McNemar's test was selected for comparing case subjects with control subjects on categorical variables. The Wilcoxon signed rank test was used for ordinal measures, such as the number of lifetime symptoms related to alcohol or drug disorders.

TABLE 1—Demographic Profile of Case Subjects (Those Admitting Physical Abuse or Neglect of Children) and Control Subjects

	Abuse		Neglect	
	Case Subjects, % (No.)	Control Subjects, % (No.)	Case Subjects, % (No.)	Control Subjects, % (No.)
Age, y[a]				
18–24	8.9 (15)	8.3 (14)	17.7 (37)	13.9 (29)
25–44	69.8 (101)	59.2 (100)	51.2 (107)	56.0 (117)
45–64	26.0 (44)	27.2 (46)	20.1 (42)	17.7 (37)
65+	5.3 (9)	5.3 (9)	11.0 (23)	12.4 (26)
Race[a]				
White	61.2 (101)	61.2 (101)	37.9 (78)	37.9 (78)
Non-White	38.5 (64)	38.8 (64)	62.1 (128)	62.1 (128)
Marital status				
Married	49.7 (84)	49.7 (84)	32.2 (67)	47.4 (99)
Separated/divorced	29.6 (50)	17.8 (30)	32.7 (68)	15.8 (33)
Widowed	4.1 (7)	19.5 (33)	9.1 (19)	18.7 (39)
Never married	16.6 (28)	13.0 (22)	26.0 (54)	18.2 (38)
Education				
Less than high school	49.1 (52)	50.9 (57)	61.4 (81)	62.0 (93)
High school graduate	50.9 (54)	49.1 (54)	38.7 (51)	38.0 (57)
Socioeconomic quartile[a]				
0–25.0	21.9 (37)	21.9 (37)	36.8 (77)	36.8 (77)
25.1–50.0	35.5 (60)	35.5 (60)	33.5 (70)	33.5 (70)
50.1–75.0	31.4 (53)	31.4 (53)	23.9 (50)	23.9 (50)
75.1–100.0	11.2 (19)	11.2 (19)	5.7 (12)	5.7 (12)
Total	44.7 (169)	44.7 (169)	55.3 (209)	55.3 (209)

[a]Variable used to match case subjects with control subjects.

TABLE 2—Proportion of Respondents with Lifetime Alcohol or Drug Disorders among Subjects Admitting Physical Abuse or Neglect of Children, Compared with Control Subjects

	Abuse			Neglect		
	Case Subjects, % (No.)	Control Subjects, % (No.)	Odds Ratio[a]	Case Subjects, % (No.)	Control Subjects, % (No.)	Odds Ratio[a]
Any alcohol or drug disorder	40.2 (68)	16.0 (27)***	5.1	56.0 (117)	16.8 (35)***	9.2
Any alcohol disorder	36.1 (60)	13.8 (23)***	4.7	45.8 (92)	15.9 (32)***	5.3
Alcohol abuse	9.0 (15)	3.6 (6)*	2.5	9.5 (19)	5.0 (10)	2.1
Alcohol dependence	27.5 (46)	10.2 (17)***	3.6	36.3 (73)	11.0 (22)***	5.2
Any drug disorder	12.5 (21)	4.8 (8)*	2.9	27.3 (56)	4.4 (9)***	10.4
Drug abuse	3.6 (6)	1.8 (3)	2.5	6.3 (13)	2.0 (4)*	4.0
Drug dependence	8.9 (15)	3.0 (5)*	3.0	21.0 (43)	2.5 (5)***	10.5

[a]Matched-pair odds ratio.
*$P \leq .05$; **$P \leq .01$; ***$P \leq .001$.

Conditional logistic regression for paired data was used to predict group membership (abuse cases vs control group; neglect cases vs control group) on the basis of the presence or absence of a drug or alcohol disorder, after controlling for depression, number in household, antisocial personality disorder, and social support.

Results

A total of 11 662 of the individuals who were successfully interviewed across the four sites did not deny being parents or acting as a caretaker for children. A total of 169 respondents (1.4%) reported abusive behavior, whereas 209 respondents (1.8%) reported neglectful behavior. Sixteen respondents (0.1%) reported both physically abusive and neglectful behavior. The inclusion of respondents reporting both abusive and neglectful behaviors did not significantly change the findings of the study. Results are presented with these individuals categorized as case subjects for both child abuse and neglect.

Table 1 presents the demographic profile of respondents admitting physical abuse or neglect. The mean age of both abusive and neglectful adults at the time of interview was 40 years. The abusive group was predominantly White (61%) with a high school education (51%) and an average socioeconomic status score of 45 (out of a possible range of 0 to 100), whereas the neglectful group consisted primarily of non-Whites (63%) with less than a high school education (61%) and a lower average socioeconomic status score (38%).

In Table 2, the results of the comparisons between case subjects and control subjects for presence of alcohol or drug disorders are presented. The lifetime prevalence of alcohol or drug disorders among subjects admitting physical abuse or neglect of children is markedly higher than that among their matched control subjects, almost without regard to type of diagnosis considered. Control subjects had prevalence rates similar to those reported among the entire sample of household respondents. Both control groups reported on average less than one alcohol or drug symptom, whereas their respective case counterparts reported substantially higher averages.

Results of the logistic regression for paired data (Table 3) give the estimated odds ratio for child physical abuse or neglect for those with a substance abuse disorder after simultaneously controlling for depressive disorder, number in household, antisocial personality disorder, and social support. Substance abuse or dependence continued to be a significant predictor of both child abuse and neglect. In fact, those adults with an alcohol or drug disorder were 2.7 times more likely to have reported abusive behavior toward children and 4.2 times more likely to have

reported neglectful behavior toward children than were their matched control subjects. Adults living in larger households or with a history of depressive disorder were also significantly more likely to have acted abusively toward children than their matched control subjects. A history of antisocial personality disorder remained a significant predictor of child neglect. Interpretation of this finding should be tempered by the consideration that child-maltreatment items are embedded in the antisocial personality disorder diagnosis, thereby increasing the possibility of finding an association between the disorder and child maltreatment. However, regression models omitting the antisocial personality disorder variable produced results similar to those reported here.

Because of concerns about a positive response set among those adults reporting high rates of symptoms, we examined the frequency of panic disorder across cases. Fewer than 5% of neglectful case subjects and 7% of abusive case subjects reported symptoms sufficient to meet criteria for panic disorder, suggesting that the high rates of substance abuse and dependence found in our study were not the result of a general positive response set. The frequency of panic disorder was slightly lower in control subjects.

TABLE 3—Odds Ratios and 95% Confidence Intervals for Child Abuse and Child Neglect and the Proportions of Case and Control Subjects

	Odds Ratio	95% Confidence Interval	% Case Subjects/ % Control Subjects
Abuse			
No. in household	1.45	1.20, 1.74	3.4/2.5[a]
Depression	4.03	1.25, 13.01	16.4/6.2
Availability of confidant	2.05	0.92, 4.59	28.1/11.0
Antisocial personality	3.36	0.88, 12.85	14.4/2.1
Substance abuse	2.74	1.20, 6.28	35.6/13.0
Neglect			
No. in household	1.11	0.94, 1.31	3.2/2.83[a]
Depression	1.80	0.60, 5.43	13.7/7.5
Availability of confidant	1.01	0.48, 2.11	29.5/15.8
Antisocial personality	18.13	2.37, 138.58	26.7/1.4
Substance abuse	4.17	1.81, 9.59	47.3/15.1

[a]Mean.

Discussion

This is the first controlled study to use a community-based sample with standardized diagnoses and assessments to explore the prevalence of alcohol or drug disorders among physically abusive and neglectful parents and caretakers. Forty percent of the adults reporting physically abusive behavior and 56% of the adults reporting neglectful behavior met criteria for an alcohol or drug disorder during their lifetime. These rates of substance abuse disorders were markedly higher than those found in a control group matched on age, race, gender, socioeconomic status, and site. The physically abusive and neglectful groups were also found to report more drug or alcohol disorder symptoms than their respective control subjects. The contribution of substance abuse disorders to predicting physical abuse or neglect remained significant after controlling for depression, household size, antisocial personality disorder, and social support.

Although the differences noted in this study are highly significant and correspond with anticipated results, it is important to underscore some limitations. First, the data are retrospective and do not include information on the temporal sequencing of drug or alcohol disorders and abuse or neglect of children. As a result, no causal inferences can be made. However, heavy users of drugs and alcohol often initiate their use in early adolescence,[23,24] before becoming parents, suggesting that substance abuse disorders are likely to be established or in process when parenting tasks are first encountered. In this study more than 25% of case parents or caretakers with substance abuse disorders began using alcohol or drugs before age 15. Additionally, it is not possible to ascertain the contribution of drugs or alcohol to *individual* acts of maltreatment. However, it is likely that the impact of drug or alcohol use at a level sufficient to warrant a formal diagnosis probably extends well beyond the immediate effects of use.

Second, the prevalence of abuse and neglect found in these data is similar to the prevalence of abuse found in a Canadian study using the Diagnostic Interview Schedule[9] and higher than the prevalence garnered from child welfare agencies. However, it is below the generally accepted prevalence of abuse and neglect in the population and is approximately equal to the estimated annual incidence rates found in other studies.[12] A number of factors may account for this result. First, only one respondent per household was interviewed, thereby underestimating the number of children abused or neglected because most children are "exposed" to more than one adult. Second, child-maltreatment behaviors are probably minimized or hidden due to the social stigma attached to the problem and are likely to be underestimated by self-report data. However, if social desirability bias or concern about self-incrimination caused some maltreating respondents to be classified as control subjects, the effect on the study should be to reduce observed effect sizes by falsely classifying some abusive or neglectful parents as control subjects. On the other hand, respondents falsely denying both child maltreatment and substance abuse disorders could increase our observed effect size.

Finally, the Diagnostic Interview Schedule does not sample the entire range of child physical-abuse and neglect behaviors. The content of the physical-abuse question suggests that parents were only asked about relatively serious occurrences. Thus, these findings could be consistent with Steele's contention that alcohol abuse is a greater factor among the more severe cases of physical maltreatment.[8]

Although this study cannot conclude a causative relationship, the use of a community sample and matched control subjects suggests that the relationship between parental substance abuse and child maltreatment is robust and cannot easily be dismissed on the basis of referral patterns, social class, or community of residence. It is also unlikely that a generic "yes" response set to psychiatric symptoms by some respondents resulted in the association, as neither case subjects nor control subjects were likely to receive a diagnosis of panic disorder, a condition

with no theoretical or empirical link to child maltreatment. Finally, the results support the hypothesis that depression, antisocial personality disorder, or lack of social support mediates some of the relationship between substance abuse and the types of maltreatment studied.

Our results suggest that parental psychopathology, independent of social ecology, is an important determinant of child maltreatment in many cases. Recognizing substance abuse in maltreating parents may alter treatment and outcomes for families in child welfare services. For example, it is widely accepted that maltreating parents with substance abuse disorders must receive treatment for their disorder before any interventions for maltreatment with the family are likely to be effective.[25]

Currently, most maltreating families do not receive a specific substance abuse evaluation,[14,26] and among known substance abusers in one study, fewer than half of the maltreating parents were referred for alcohol or drug treatment.[14] Data from one state reveal that substance abuse was identified as a problem in only 8% of substantiated child abuse or neglect cases, and treatment was provided for only 4%. Child welfare systems will need to carefully examine their child maltreatment case-finding mechanisms and their substance abuse evaluation and treatment programs. □

Acknowledgments

This work was supported in part by NIMH grant P50 MH48197.

The authors would like to thank Drs Brenda Booth and Joel Milner for their comments.

References

1. Daro D, McCurdy K. *Current Trends in Child Abuse Reporting and Fatalities: The Results of the 1990 Annual Fifty State Survey.* Chicago, Ill: National Committee for Prevention of Child Abuse; 1991.
2. Black R, Mayer J. Parents with special problems: alcoholism and opiate addiction. *Child Abuse Negl.* 1980;4:45–54.
3. Besharov DJ. The children of crack: will we protect them? *Public Welfare.* Fall 1989;47: 6–11.
4. Pelton L. Child abuse and neglect: the myth of classlessness. *Am J Orthopsychiatry.* 1978;48:608–617.
5. Blazer D, Crowell BA, George LK. Alcohol abuse and dependence in the rural south. *Arch Gen Psychiatry.* 1987;44:736–740.
6. Parker S, Greer S, Zuckerman B. Double jeopardy: the impact of poverty on early child development. *Pediatr Clin North Am.* 1988;35:1227–1241.
7. *Study Findings: Study of National Incidence and Prevalence of Child Abuse and Neglect: 1988.* Washington, DC: US Dept of Health and Human Services, National Institute of Child Health and Human Development, National Center on Child Abuse and Neglect; 1988.
8. Steele B: Psychodynamic factors in child abuse. In: Helfer R, Kempe C, eds. *The Battered Child.* Chicago, Ill: University of Chicago Press; 1987:81–114.
9. Bland R, Orn H. Psychiatric disorders, spouse abuse and child abuse. *Acta Psychiatr Belg.* 1986;86:444–449.
10. Smith S, Hanson R, Noble S. Parents of battered babies: a controlled study. *Br Med J.* 1973;4:388–391.
11. Smith S. *The Battered Child Syndrome.* London, England: Butterworth; 1975.
12. Gelles RJ, Cornell CP. *Intimate Violence in Families.* 2nd ed. Newberry Park, Calif: Sage; 1990.
13. Milner JS, Chilamkurti C. Physical child abuse perpetrator characteristics. A review of the literature. *J Interpersonal Violence.* 1991;6:345–366.
14. Kameen MC, Thompson DL. Substance abuse and child abuse-neglect: implications for direct-service providers. *Personnel Guidance J.* January 1983;61:269–273.
15. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders. 3rd ed, revised.* Washington, DC: American Psychiatric Association; 1987.
16. Leonard K, Jacob T. Alcohol, alcoholism, and family violence. In: VanHasselt V, Morrison R, Bellack A, Hersen M, eds. *Handbook of Family Violence.* New York, NY: Plenum Press; 1988:383–406.
17. Eaton WW, Anthony JC, Tepper S, Dryman A. Psychopathology and attrition in the epidemiologic catchment area surveys. *Am J Epidemiol.* 1992;135:1051–1059.
18. Robins L, Helzer J, Croughan J, Ratcliff K. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Arch Gen Psychiatry.* 1981;38:381–389.
19. Anthony J, Folstein M, Romanoski A, et al. Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis. *Arch Gen Psychiatry.* 1984;42:667–675.
20. Zuravin S. Child abuse, child neglect, and maternal depression: is there a connection? In: *Research Symposium on Child Neglect.* Washington, DC: US Dept of Health and Human Services; 1988.
21. Polansky N, Ammons P, Gaudin J. Loneliness and isolation in child neglect. *Social Casework.* 1985;66:38–47.
22. Robins L. *Deviant Children Grown Up: A Sociological and Psychiatric Study of Sociopathic Personality.* Baltimore, Md: Williams & Wilkins; 1966.
23. Windle M. Alcohol use and abuse: some findings from the National Adolescent Student Health Survey. *Alcohol Health and Research World.* 1991;15:5–10.
24. Klitzner M, Fisher D, Stewart K, Gilbert S. *Substance Abuse: Early Intervention for Adolescence.* Princeton, NJ: Robert Wood Johnson Foundation; 1992:4–6.
25. Kinscherff R, Kelley S. Substance abuse: intervening with substance abusing families. *The APSAC Advisor.* 1991;4:3–5.
26. *Third Special Report to the U.S. Congress on Alcohol and Health.* Rockville, Md: National Institute on Alcohol Abuse and Alcoholism; 1978.

Consequences of the North American Free Trade Agreement for Health Services: A Perspective from Mexico

ABSTRACT

Objectives. The purposes of the study were to assess the potential impact of the North American Free Trade Agreement (NAFTA) on medical care in Mexico and to identify internal measures Mexico could take to increase the benefits and minimize the risks of free trade.

Methods. The dual nature of the health sector is examined; the Mexican, Canadian, and US health care systems are compared; and modes and consequences of international exchange of health services are analyzed.

Results. Four issues require immediate attention: accreditation of health care facilities, licensing and certification of professionals, technology assessment, and financial equity.

Conclusions. NAFTA offers opportunities for positive developments in Mexico, provided risks can be anticipated and preventive measures can be taken to avoid negative impacts on the health system. Medical services, like other elements of the Mexican economy, must be modernized to respond to the demands of global competition. The Mexican National Academy of Medicine has recommended to the Mexican government (1) internal strengthening of the Mexican health care system to improve its ability to respond to the new conditions created by NAFTA and (2) a gradual process to facilitate equitable and mutually beneficial interactions among the three countries. (*Am J Public Health.* 1994;84: 1591–1597)

Julio Frenk, MD, PhD, MPH, Octavio Gómez-Dantés, MD, MPH, Carlos Cruz, MHE, Fernando Chacón, MPA, Patricia Hernández, MD, MPH, and Phyllis Freeman, JD

Introduction

After a tense and intense debate, the North American Free Trade Agreement (NAFTA) gained formal approval for implementation beginning January 1, 1994. In the midst of the discussion about the agreement's immediate economic effects, its long-term impact on social sectors such as health received far less attention. In Mexico the NAFTA negotiation process, initiated 2 years ago, did trigger a policy debate on the future of the health system. In this article we discuss why NAFTA could have far-reaching effects on personal health services in Mexico, identify the nature of those effects, discuss their potential risks and benefits to society, and summarize the recommendations offered by the Mexican National Academy of Medicine in 1992 to the Ministry of Trade and Industrial Development, the agency leading the NAFTA negotiations for Mexico.[1]

Economic integration and trade liberalization are bound to have profound effects on every aspect of social life in Mexico. In health, NAFTA will influence five major areas: environmental and occupational health; regulation of food, drugs, water, and sanitation; capital investments in facilities and development and sales of medical equipment, devices, pharmaceutical products, and other such "inputs" for health care; training and utilization of personnel; and regulation, financing, and delivery of health services.

The first three of these areas are clearly linked not only to the health of the population but also to overall economic growth. Therefore, they were an integral part of discussions about NAFTA since the early stages. In contrast, the last two areas were subsumed under a broader negotiation group dealing with the entire service sector of the economy.

To help fill the resulting analytic gap, we focus on the implications of NAFTA for the regulation, financing, and delivery of personal health services. The central point of this paper is that, just like the rest of the economy, health services in Mexico must be modernized to respond to the demands of global competition. This process of modernization makes issues such as accreditation of health care units, licensing and certification of professionals, technology assessment, and financial equity important topics for examination and discussion. Attention to these issues can strengthen the capacity of the state to define and enforce standards to promote quality, protect consumers, and help guarantee basic equal opportunities for all Mexicans.

In the first part of the paper we discuss the characteristics of health and health care that must guide the analysis of the consequences of NAFTA. In the second part we compare the Mexican, Canadian, and American health systems. We then move to an examination of the effects of NAFTA on the main modes of

Julio Frenk is with the Mexican Health Foundation and the Faculty of Medicine, National Autonomous University of Mexico, Mexico City. Octavio Gómez-Dantés, Carlos Cruz, Fernando Chacón, and Patricia Hernández are with the National Institute of Public Health, Cuernavaca, Mexico. Phyllis Freeman is with the Law Center and the John W. McCormack Institute of Public Affairs, University of Massachusetts, Boston.

Requests for reprints should be sent to Julio Frenk, MD, PhD, MPH, Fundación Mexicana para la Salud, Periférico Sur No. 4809, Col. El Arenal Tepepan, Delegación Tlalpan, 14610 México, DF, Mexico.

This paper was accepted March 2, 1994.

TABLE 1—Basic Health Indicators for Mexico, Canada, and the United States, circa 1990

	Mexico	Canada	United States
Population (1×10^6)	81.1[a]	26.5[g]	249.2[g]
Overall fertility rate	3.2[b]	1.7[f]	1.8[f]
Life expectancy at birth, y			
Men	66.4[b]	73.4[h]	71.2[g]
Women	72.9[b]	79.8[h]	78.2[g]
Infant mortality rate	38.1[b]	7.3[h]	10.1[h]
% gross national product spent on health	4.0[c]	8.6[h]	12.2[h]
Health expenditure (1×10^3 dollars)	5.1[d]	50.4[g]	662.2[h]
Hospital beds per 1000 inhabitants	1.3[e]	6.8[f]	4.4[f]
Physicians (1×10^3)	130.0[f]	57.4[f]	585.6[f]

Sources.

[a]*Resultados preliminares del XI Censo General de Población y Vivienda*. Mexico City, Mexico: Instituto Nacional de Estadística, Geografía e Informática; 1991.

[b]*Programa Nacional de Población 1989–1994*. Mexico City, Mexico: Consejo Nacional de Población; 1991.

[c]*Programa Nacional de Salud 1990–1994*. Mexico City, Mexico: Secretaría de Salud; 1990:18.

[d]Includes only public sector health and social security expenditures. Fundación Mexicana para la Salud, Fundación Javier Barros Sierra, Carnegie Corporation of New York. *Perfiles de la Salud hacia el año 2010*. Mexico City, Mexico: Fundación Mexicana para la Salud; 1988. (6 553 466 million 1987 pesos at 1249 pesos/dollar = 5 064 502 319 dollars.)

[e]Estimated figure. Sistema Nacional de Salud. *Boletín de Información Estadística* (Mexico). 1991 (10).

[f]*Las Condiciones de Salud en las Américas*. Washington, DC: Pan American Health Organization; 1990.

[g]*World Health Statistics*. Geneva, Switzerland: World Health Organization; 1991.

[h]Long H. Health care in North America: resource allocation in Canada, the United States, and Mexico. Presented at the Seminar on Rationing and Use of Technology in Health in Mexico; November 4–5, 1991; Mexico City, Mexico.

international exchange of health services and end with a series of recommendations. Although our analysis centers on Mexico, we hope it will stimulate more dialogue with those familiar with the Canadian and American points of view.

The Dual Nature of Health Services

Understanding the special character of health services is a prerequisite to any analysis of their relationship to trade liberalization. Unlike other products, health services have a dual nature. On the one hand they represent a fundamental component of social development and well-being; on the other, a growing sector of the economy.

In modern times, the quest for health has been strengthened by unprecedented advances in science and technology. Although the health status of a population is fundamentally determined by its living conditions, personal health services can substantially change the incidence and severity of disease.

The fact that in this century health services have become effective has given them a positive value and made them the object of increasing demand by society. Thus, access to such services has become an indicator of the quality of life in a community. Health is not only one of the ultimate aims of economic growth; it is also one of its conditions, because it is a determinant of productivity. In addition, the provision of health services has been used as a means for redistributing wealth and, together with education, as a fundamental element in the promotion of equal opportunities. Because of their connection with the essential processes of life—birth and death, enjoyment and pain—health services are at the center of solidarity. Access proportional to need—not to privilege or ability to pay—has become a fundamental component of equity and social justice.

At the same time, the financing and delivery of services has become a dynamic sector in all modern economies.[2] It is an important source of employment, a generator of demand for all kinds of inputs, and a driving force for technological innovation. As such, it has fueled the emergence of a vast medical-industrial complex. However, the market for medical services is far from perfect. Among its most important imperfections are externalities, uncertainty, consumer ignorance, induced demand, and monopolistic conditions.[3] Together with the ethical aspects of medical care and the role of health in social well-being, these market imperfections mean that a collective mediator between producers and consumers—which in all modern societies is the state—is required in order to obtain an optimal distribution of resources, protect the interests of the community, and ensure maximum social benefit from health services.

The complex subject of state intervention in the field of health goes beyond the scope of this paper. Suffice it to say that the state can carry out three fundamental activities related to personal health services: regulation, financing, and direct delivery.[4] Regulation has always been a public responsibility, but there has been substantial variation in government control over financing and delivery across different countries and at different moments in history. For this reason, analysis of the effects of NAFTA requires that we first examine the structure and function of the health systems of Canada, Mexico, and the United States, with emphasis on these three activities.

Comparative Analysis

Good descriptions of the health systems of Canada, Mexico, and the United States already exist in the literature.[5–7] Because many North American readers are likely to be less familiar with the Mexican situation than with health conditions and services in Canada or in the United States, and because Mexico is dissimilar to Canada and the United States in ways that become even more important under trade liberalization, we provide a discussion of the salient features of the Mexican health system. This discussion supplements the information presented in Table 1, which presents various indicators of health conditions, expenditures, and resources in the three NAFTA countries, and Table 2, which shows the main causes of death in each country.

Because the Mexican Constitution has explicitly recognized a right to health protection, it assigns a key role to the state in conducting national health policy. The National Health System is made up of three components. The first includes those governmental organizations providing services for the uninsured population, principally the rural and urban poor. The main institutions involved are the Ministry of Health, the National Institutes of Health, the IMSS–Solidarity Program, the Health Services of the Federal District,

TABLE 2—Main Causes of Death in Mexico, Canada, and the United States, 1991

	Deaths per 100 000
Mexico	
Cardiovascular diseases	57.9
Accidents	50.3
Cancer	47.6
Infectious intestinal diseases	31.0
Canada	
Cardiovascular diseases	302.4
Cancer	195.8
Cerebrovascular diseases	53.0
Accidents	35.3
United States	
Cardiovascular diseases	336.0
Cancer	183.9
Cerebrovascular diseases	75.1
Accidents	46.7

Source. Reproduced with permission from *World Health Statistics Annual 1991.* Geneva, Switzerland: World Health Organization; 1992.

TABLE 3—Main Features of the Health Systems of Mexico, Canada, and the United States

	Mexico	Canada	United States
Regulation			
Registration of facilities	+	+++	+++
Certification of units	+	+++	+++
Technology evaluation	+	+++	+++
Cost control	+	+++	++
Private fee control	+	+++	+
Certification of schools	+	+++	+++
Certification of degrees	+	++	+++
Continuing education	Voluntary	Compulsory	Recertification
Validation of licenses	National	Provincial	State
Practicing requirements			
General medicine	Registration	License	License
Specialty	Certificate	Certificate	Certificate
Power of professional associations	+	+++	+++
Financing			
Main financial source	Public	Public	Private
Local financing	+	+++	+++
Social security	++	+++	++
Private insurance schemes	+	+	+++
Delivery of health care			
Coverage	Incomplete	Universal	Incomplete
Main type of services	Curative	Curative	Curative
Facilities' ownership	Public	Public and private	Private
Institutional organization	Fragmented	Integrated	Fragmented
Decentralization	+	+++	+++
Freedom of choice	+	++	+++
Ability to sue	+	++	+++

Note. + = low; ++ = medium; +++ = high.

and an agency in charge of family welfare. The second component, social security, comprises an institute covering workers in the formal private sector of the economy (Instituto Mexicano del Seguro Social [IMSS]) plus separate organizations for federal civil servants, some local government employees, the armed forces, and the employees of the national oil company. In contrast to the European experience, social security in Mexico has developed its own network of hospitals and salaried personnel, thereby combining the financing and delivery of services in a single agency. The third component, the private sector, is made up of a collection of health care providers working in hospitals, ambulatory clinics, offices, and indigenous (folk) medicine on a for-profit basis.

Almost 55% of the population subscribe to a social security institution, 35% receive care from agencies for the uninsured, 4% always use private services, and 6% have no easy access to permanent facilities (although they may benefit from special campaigns, such as vaccination campaigns, which have achieved virtually universal coverage). Revenue for public institutions is drawn from three sources: (1) the federal budget; (2) employee and federal contributions that finance social security benefits for civil servants; and (3) employees and employers in the private sector and a federal subsidy for the social security institute covering private sector workers. There is little voluntary medical insurance; instead, a large proportion of payments to private providers are made directly by the patients.

The government has major responsibility for financing, for providing services, and for regulating health-related activities. However, this last function has often been limited to issuing rules, while enforcement has remained weak. Responsibility for licensing of health professionals falls on institutions of higher education rather than on the National Health Service. Current procedures require that degrees be endorsed by the Ministry of Public Education. The National Academy of Medicine, a private nonprofit organization, coordinates the certification of specialists. Absence of certification of a physician in a specialty imposes no legal restriction on his or her practicing in that or any other specialized field of medicine.

When the Mexican system is compared with those of its North American neighbors, profound imbalances among the three countries become apparent (Table 3). Each system has its own strengths and weaknesses. Mexico has the benefit of a constitutional right to health protection and a recognition of health as one of the main components of social policy.[8] Health institutions currently extend care to almost the entire population. The Mexican system's principal weakness is its segmented nature, which results in large gaps in the quality of care provided to different population groups. The strength of the Canadian system is its universal coverage with a high degree of equity and decentralized control that has preserved consumer choice of primary provider. Recently, however, this system has been facing cost control problems. The American system is outstanding for its scientific and technological development but is confronting a severe crisis

whose central elements are incomplete coverage of services; fragmentation of the system, with striking inequalities; an escalating rise in costs; excessive use of technology; and an expensive combination of legal complaints and defensive medical practice.

Beyond the characteristics of each system, a comparative analysis reveals profound imbalances. These include

- Imbalances in health conditions, including major differences in mortality, morbidity, and fertility as well as in patterns of epidemiological transition.
- Imbalances in resources devoted to health care. Mexico spends, in absolute terms, 1/10 the amount spent by Canada and 1/30 the amount spent by the United States.
- Imbalances in the regulation of health services. Mexico lags behind Canada and the United States considerably in accreditation of health care facilities, certification of practitioners, and technology evaluation.
- Imbalances in the dominant forms of financing. In Mexico, these are mostly public and centralized; in Canada, they are public and decentralized; in the United States, the complex mixture includes decentralized private insurance, public centralized insurance (Medicare), and insurance sponsored by the state and federal governments (Medicaid).
- Imbalances in the forms of private sector participation. In Mexico participation is fragmented and is financed mainly by direct payment from the patient, whereas in the United States private medical insurance is the dominant form of financing.
- Imbalances in the power of the patient. Whereas in the United States patients may file excessive lawsuits, in Mexico patients with low purchasing power have little recourse if they are dissatisfied with the quality of services rendered.
- Finally, one feature that distinguishes the Canadian and American systems from the Mexican system is the separation between financing and delivery of services (although health maintenance organizations tend to bring both functions together). In Canada the combination of equitable financing with predetermined expenditure limits appears to account for better health system performance than in the United States.

One of the few features common to the three systems is the predominance of curative medicine.

A primary and paradoxical conclusion based on this analysis is that Mexico is the NAFTA country with the highest degree of direct government involvement in the financing and delivery of health services but the weakest capacity to regulate the private medical care market.

Given the marked asymmetries among the health systems of the three countries, what is the likely impact of trade liberalization on Mexico, the least developed link in this commercial chain?

The European Experience

In attempting to answer this question, it is useful to consider the European Economic Community (EEC) as a precedent. The fundamental tenets of NAFTA differ from those of the EEC. NAFTA is a limited treaty involving only the movement of goods, services, and capital. It does not imply the level of integration of the EEC. Additionally, the differences between Mexico and its two neighbors are huge in comparison with the differences between poor and rich countries of the EEC. As Schlefer states:

> NAFTA is an experiment in opening the economies of industrialized nations to a major Third World country. Japan has no such agreement with its East Asian trading partners, and although the European Community admitted Spain, Portugal, and Greece into its memberships, they are not comparable to Mexico. Spain has a per capita income close to half that of the U.S., while Mexico's is a tenth as high. Portugal and Greece are poorer than Spain but still have per capita incomes twice Mexico's, and they contain only six percent of the total population of the EC; Mexico contains almost a quarter of North America's population. The EC is bringing poorer member countries up to Community standards with billions of dollars of development assistance and the enforcement of common environmental and workplace regulations. . . . In contrast, NAFTA proposes that the U.S. and Canada take Mexico as is, while Mexico pays back a debt close to $75 billion—a kind of Marshall Plan in reverse.[9(p114)]

In contrast to NAFTA, the EEC has been firmly committed to an integrated social policy intent on defining and managing environmental, health, labor, and social relations in each of the member countries. The EEC agreement establishes the right to free movement in EEC countries for all EEC workers. This freedom of movement is accompanied by the guarantee of social security eligibility in whichever member state a worker resides.[10] The EEC's social charter, established in 1961, specifies the right to social security and medical coverage, occupational safety, decent remuneration, membership in national and international labor organizations, collective bargaining, and vocational training. It also provides for the social and economic protection of children and pregnant women.[11]

Modes of Interaction

Even before NAFTA was ratified, trade liberalization was producing changes in Mexican society. It is unlikely that the National Health Service will remain immune to such changes. Among other effects, NAFTA will lead to a more intense exchange of personal health services, especially between the United States and Mexico. Below we discuss five major modes of exchange: movement of service users, movement of service providers, establishment of health care units in other countries, exportation of services, and international joint projects.

Movement of Service Users

This phenomenon is already prevalent along the border between Mexico and the United States. Trade liberalization caused by the creation of free zones and the expansion of the *maquiladora* industry has generated a large demand for health services along this border. The patterns identified are likely to intensify, and they portend national effects under NAFTA.

Two problems arise as a result of this growing interaction. First, inadequate formal channels of communication, referral, and follow-up between providers on different sides of the border are likely to have a negative impact on the continuity, quality, and costs of care. Second, the absence of financing mechanisms to pay for health services regularly received across the border (nonportability of insurance) has repercussions on accessibility, equity, and efficiency.

Movement of Health Service Providers

Under NAFTA, health service providers may eventually cross borders to deliver services in another country. A potential advantage is improvement in the cultural match between consumers and providers for citizens who have migrated (i.e., Mexican health professionals could serve Mexican Americans in the Southwestern United States and US health professionals could serve American retirees in Mexico). Such exchanges could

improve the transfer of knowledge and clinical procedures. For health professionals to practice in these neighboring countries, licensing and certification criteria in the NAFTA countries must be made consistent and interchangeable.

Establishment of Health Care Units in Other Countries

There are few examples of foreign medical care units established in the North American region. Presently there is no legal restriction on investment in this area in Mexico, yet it took a structural adjustment program and NAFTA to create a more entrepreneurial atmosphere that is likely to lead to the development of a foreign health care industry in Mexico. Many local providers may face considerable disadvantages compared with foreign corporations, owing to preferences of the upper and middle classes in Mexico and to such problems in the Mexican private sector as outdated technology, uneven quality of medical care, and limited financial capacity for growth. Another factor that might facilitate the establishment of American corporations in Mexico is the lax character of Mexican legislation regarding the installation of health care units, whether national or foreign.

Nevertheless, foreign medical care units should not represent a threat to the National Health Service. It is doubtful that immediate dramatic changes will result from NAFTA. Therefore, Mexican organizations will have time to adapt. In addition, trade liberalization offers potential benefits. Foreign health care units can satisfy unmet needs, create jobs, decrease capital flight by reducing the number of Mexican consumers seeking health care abroad, promote competition beneficial to consumers, and introduce efficient models of service delivery and managerial technology (cost accounting, prepayment systems, etc.).

To achieve these potential benefits, Mexico must have regulation sufficient to avoid intrusion of the sorts of inequities and inefficiencies characteristic of the American system. In particular, it will be imperative to deter patient discrimination and unnecessary procedures. For the establishment of foreign units to have a positive impact on the existing system, Mexico will require effective means of ensuring equity, quality, and cost containment. Experience shows that such measures are best implemented before the market opens. Otherwise, vested interests are likely to increase the political and economic costs of introducing reforms.

At the same time, trade liberalization creates opportunities for the establishment of Mexican health care units in the United States. The high costs of medical care and the insufficient coverage of many Hispanic populations offer Mexican providers the opportunity to address the needs of a significant market.

Export of Services

The exchange of diagnostic or therapeutic information among institutions providing health care in different countries amounts to export of services, often by telecommunication. Although it holds much potential for the future, this mode of interaction has the least likelihood of implementation in the short term, because it will depend on substantial communication infrastructure and standardization of referral patterns.

Cooperation in International Joint Projects

Another interesting opportunity offered by NAFTA involves the development of international consortia. Joint projects could introduce innovations in the delivery and financing of health services. Equally attractive are proposals for linking programs of scientific research and higher education for mutual benefit. Some cooperative projects already exist in the region, and NAFTA will provide the impetus for more.

Consequences of Trade Liberalization

The NAFTA text contains minimal references to personal health services. Most of the sections devoted to health relate to sanitary regulation and to occupational and environmental health. However, Annex 1210 includes a provision that is relevant to medical services. This provision indicates that "relevant bodies" will negotiate "mutually acceptable professional standards and criteria for licensure and certification of professional service providers and [will] provide recommendations on mutual recognition."[12]

Beyond this specific provision, the five modes of interaction are likely to create opportunities and risks for Mexico that must be borne in mind in order to optimize potential benefits and minimize risks. Table 4 summarizes some possible consequences of NAFTA; it contrasts the prospects in Mexico with those for the United States and Canada, although obviously there are great differences between the latter two.

As suggested by Table 4, NAFTA will create the possibility of expanding the market for the medical-industrial complex in Mexico. It is likely to promote national and foreign investment, which in turn will create jobs and improve the technological content, competitiveness, accessibility, quality, and productivity of health services in Mexico.

The major risks relate to a greater dominance of curative medicine, a lack of consumer protection from unqualified providers, possible brain drain, an increase in the practice of defensive medicine, cost inflation, deeper inequalities, and underutilization of the medical infrastructure already in Mexico.

Whereas taking advantage of the opportunities will require various structural changes and trade strategies, minimizing the risks does not imply large direct costs but does require political will.

Recommendations

A variety of recommendations consistent with the preceding analysis were offered by the Mexican National Academy of Medicine to the Mexican government in December 1992. A summary of the advice follows. One set of recommendations is directed to internal strengthening of the National Health Service with the objective of increasing its ability to respond to the new conditions. A second set is intended to deal with the exchange of services among the three countries.

Internal Strengthening of the Health System

Despite the asymmetry of health care resources among the three countries, Mexico has been able to develop an ample infrastructure. Hence, the first priority is protection of progress already achieved in institutional development, training of the health services work force, and extension of coverage. At the same time, it is necessary to identify the structural aspects of the National Health Service that may complicate the pursuit of opportunities offered by NAFTA and the reduction of NAFTA-related risks.

Like the rest of the economy, health services in Mexico must be modernized to respond to the demands of global competition. Given market imperfections, strengthening the capacity of the state to define and enforce standards that will promote quality and protect consumers

TABLE 4—Opportunities and Risks of the North American Free Trade Agreement for Mexico in the Health Care Areas

Mode of Interaction	Opportunities		Risks	
	Direct	Indirect	Direct	Indirect
Movement of users				
Americans and Canadians to Mexico	Market expansion Economies of scale Growth in the medical-industrial complex Currency earnings Establishment of referral systems Greater freedom of choice for users	Incentives for improvement in quality Incentives for investment Growth of the private sector and insurance Reduction of unemployment and underemployment in the medical sector Creation of jobs in complementary economic sectors	Discrimination against Mexican patients Increase in care costs More lawsuits against Mexican providers Spreading of infectious diseases	Increase in profit-oriented medicine Problems in disease control and surveillance
Mexicans to the United States and Canada	Access to specialized medical procedures and technology	Satisfaction of needs not met in Mexico	Underutilization of the medical infrastructure in Mexico	Fall in investment Capital flight Reduction in local revenue
Movement of providers				
Americans and Canadians to Mexico	Greater choice for Mexican users Complementarity of medical services Improvement of Mexican procedures for professional licensing to bring them in line with international standards Increase in competition Better quality of care	Strengthening of Mexican regulatory bodies Wider market for Mexican medical schools and faculties to train Americans and Canadians Better quality of health professional education	Increase in unemployment and underemployment of Mexican physicians Strengthening of predominance of curative medicine Increase in costs Selective movement of the least competent providers	With inadequate agencies, lack of protection for Mexican consumers With inflexible rules, defensive medicine Resurgence of inequality Capital flight due to currency remittances
Mexicans to the United States and Canada	Satisfaction of Hispanic demand for care	Improved continuing education for Mexican providers	Lack of interest in Mexican market Brain drain	Increase in costs
Establishment of units				
American and Canadian units in Mexico	Increase in foreign investment Satisfaction of demand for health services by population segments with ability to pay Technology and knowledge transfer	Creation of jobs Greater competition Development of private medical insurance Renewal and expansion of medical-industrial complex Greater coverage	Fall in investment in local capital Deeper inequities Unnecessary medical treatment Transfer of inefficiencies Transfer of environmental and occupational risks	Financial crisis for Mexican private institutions due to fall in demand Fall in public responsibility for satisfaction of social needs
Mexican units in the United States and Canada	Access to Hispanic market	Solidarity with Hispanic groups and legal and illegal Mexican seasonal workers	...	...
Services				
Importation	Adoption and improvement of international norms and standards	Improvement of communications structure	Disincentives to scientific and technological research and development Reduction of local investment Capital flight	Worsening of foreign dependence Increase in costs
Exportation	Currency earnings Vertical integration of medical-industrial complex	Incentives to investment Creation of jobs Scientific and technological development	Lack of interest in Mexican market	Increase in costs
Joint projects	Establishment of formal means of interaction Exchange of information Better educational opportunities Complementarity of services Increased mutual understanding	Increase in financial resources for research and education Optimization of resources Strengthening of research and education programs Enhanced common security	...	...

merits the highest priority. To this end, the establishment of new health care facilities in Mexico should conform to rules that are explicit, not ambiguous as they have been to date. We do not suggest that this process should be centralized. On the contrary, at least part of the accreditation authority should be decentralized to Mexico's constituent states, while maintaining minimum quality standards for the entire nation. In addition, the formulation of new standards should involve all interested parties. This broad social participation should concentrate on the design of standards while recognizing that the government is responsible for ensuring compliance.

Another area in need of attention is licensure and certification of professionals. Mexico already has a structure for certifying specialists that is similar to those in the United States and Canada. The most important differences between Mexico and the other two countries lie in the licensing of general practitioners and other professionals with a first degree. As is the case in England, medical education sufficient to begin practice in Mexico occurs at the undergraduate level, not as graduate education. Therefore, it is necessary to conduct studies about the best way in which licensing could actually allow verification of the level of competence.

In addition to personnel, technology plays a fundamental role in the quality of care. An enduring mechanism must be established to evaluate rapidly appearing new technologies. The aim is to generate information on the efficacy, safety, and cost-effectiveness of technology to serve as an objective basis for decision making.

Finally, equity of the system must be addressed. Regardless of the configuration of financing and delivery, the state must ensure that no one is excluded from the benefits of health care and that access is proportional to health needs and not to purchasing power.

The legal framework and procedures for the protection of users of health services in Mexico also require modernization. Consumer protection must be strengthened without absorbing resources vital to the continued development of the health services system. To accomplish this balance will require the establishment of a participatory scheme whereby most cases can be resolved at the level of professional disciplinary bodies.

In general, regulatory efforts must seek a balance. A framework of rules that is too lax imposes costs on patients and society, but excesses in regulation also have undesirable side effects.

Proposals for the International Exchange of Services

Undoubtedly, opening the market will produce a profound transformation in the composition of the Mexican National Health System. Therefore, the main recommendation for the NAFTA negotiation process was for Mexico to adopt a position of gradual change. Indeed, time is needed to implement the internal measures suggested above, which are indispensable for the protection of Mexican consumers. Nevertheless, it is possible to formulate short- and middle-term proposals.

Measures worthy of pursuit during the early phase of NAFTA implementation include the following:

1. Reciprocal recognition of private health insurance among the three countries.

2. Payment by public and private American and Canadian insurance agencies for services rendered in Mexico by providers that follow international certification procedures. In particular, Medicare could cover health care expenses of the retired American community in Mexico, which amounts to 12% of social security recipients living abroad.[13]

3. Establishment in Mexico of low-technology facilities—hospitals for the chronically ill, convalescence hospitals, nursing homes—to care for American and Canadian patients.

4. Establishment in the United States and Canada of Mexican medical care facilities for the Hispanic populations living in these countries, as well as for seasonal workers from Mexico and other Latin American countries.

5. Adoption of compatible information systems to speed patient referral, especially in the US–Mexico border area.

Middle-term proposals include a study of the feasibility of Mexican agencies' financing health services provided in the United States and Canada when the patient's needs so require and the establishment of a tripartite group to analyze ways of strengthening linking mechanisms in health care, with the ultimate objective of obtaining the greatest possible benefit for the populations of the three countries.

In conclusion, Mexico faces a process of change offering enormous development potential for health services in terms of improvement, diversification, and expansion. To take full advantage of such opportunities, Mexico must anticipate and address the risks entailed in the process. This will require both analytical capacity and the will to reach mutually beneficial solutions on the part of all interested parties. □

Acknowledgments

This paper is based, with modifications, on a summary of a document that was prepared by the authors as members of the technical secretariat for the Working Group on NAFTA and Medicine established by the National Academy of Medicine of Mexico. We thank the members of this group and particularly Adolfo Martínez-Palomo, who inspired and facilitated our work, first as vice-president and later as president of the Academy. We also acknowledge the participation of the other members of the technical secretariat, Luis Felipe Abreu and Celia Pérez. Responsibility for the content of this paper rests solely with the authors.

References

1. *El Tratado de Libre Comercio y los Servicios Médicos: Elementos para el Análisis y la Negociación.* Mexico City, Mexico: Academia Nacional de Medicina; 1992.
2. Klarman HE. *The Economics of Health.* New York, NY: Columbia University Press; 1965.
3. Cullis JG, West PA. *The Economics of Health: An Introduction.* London, England: Martin Robertson; 1979.
4. Frenk J, Donabedian A. State intervention in medical care: types, trends and variables. *Health Policy Plann.* 1987;2:17–31.
5. Iglehart JK. Canada's health care system. *N Engl J Med.* 1986;315:202–208, 778–784, 1623–1628.
6. Frenk J. The political economy of medical underemployment in Mexico: corporativism, economic crisis and reform. *Health Policy.* 1990;15:143–162.
7. Wilson FA, Neuhauser D. *Health Services in the United States.* 2nd ed. Cambridge, Mass: Ballinger Publishing Co; 1987.
8. Soberón G, Frenk J, Sepúlveda J. The health care reform in Mexico: before and after the 1985 earthquake. *Am J Public Health.* 1986;76:673–680.
9. Schlefer J. What price economic growth? *Atlantic Monthly.* December 1992:113–118.
10. Minet P. *Full Text of the Rome Treaty and an ABC of the Common Market.* London, England: Christopher Johnson Publishers; 1962.
11. McGuiness M. Hacia una política social en América del Norte: ensayo sobre la importancia de la carta social y el fondo estructural europeos para el TLCAN. In: Vega-Cánovas G, ed. *Mexico, Estados Unidos y Canadá 1991–1992.* Mexico City, Mexico: El Colegio de México; 1994.
12. *Tratado de Libre Comercio para América del Norte.* Mexico City, Mexico: Secretaría de Comercio y Fomento Industrial; 1992.
13. Warner DC. *Mexican Provision of Health and Human Services to American Citizens: Barriers and Opportunities.* Washington, DC: Commission for the Study of International Migration and Cooperative Economic Development; 1990.

Stage of Cancer at Diagnosis for Medicare HMO and Fee-for-Service Enrollees

Gerald F. Riley, MSPH, Arnold L. Potosky, MHS, James D. Lubitz, MPH, and Martin L. Brown, PhD

ABSTRACT

Objectives. Health maintenance organizations (HMOs) with Medicare contracts often provide cancer screening and preventive services not covered under fee-for-service. This study compared cancer patients in HMOs and fee-for-service on stage at diagnosis.

Methods. The study examined stage at diagnosis for aged Medicare enrollees in HMOs and fee-for-service, using information from the Surveillance, Epidemiology, and End Results program, linked with Medicare enrollment files. Twelve cancer sites were investigated, and demographics, area of residence, year of diagnosis (1985 to 1989), and education at the census tract level were controlled.

Results. HMO enrollees were diagnosed at earlier stages for cancers of the female breast, cervix, colon, and melanomas and at later stages for stomach cancer. There were no differences for cancers of the prostate, rectum, buccal cavity and pharynx, bladder, uterus, kidney, and ovary. HMO effects were strongest in areas with large, mature HMOs.

Conclusions. Compared with fee-for-service enrollees, HMO enrollees were diagnosed at earlier stages for cancer sites for which effective screening services are available. The earlier detection of certain cancers among HMO enrollees may result from coverage of screening services and, perhaps, promotion by HMOs of such services. (*Am J Public Health.* 1994;84:1598–1604)

Introduction

Because of their cost-containment potential, health maintenance organizations (HMOs) and other managed care delivery systems are cornerstones of many health reform proposals. Consequently, there is considerable interest in how access to and quality of care in HMOs compares with that in fee-for-service. Access and quality are prominent issues in the Clinton Administration's Health Security Act, which calls for evaluation of care provided by individual health plans and for feedback to consumers. Most studies to date have found few differences between HMOs and fee-for-service on access or quality.[1]

Some studies have examined stage of cancer at diagnosis to compare care in HMO and fee-for-service settings because it is strongly associated with survival. These studies have generally found no difference between HMO and fee-for-service in stage at cancer diagnosis.[1–5] With the exception of Brown et al.,[1] these studies have been limited to single HMOs, and most have been limited to one cancer site.

Our study examined differences in stage of cancer at diagnosis between aged Medicare beneficiaries enrolled in HMOs and those treated in fee-for-service. Staff model, group model, and Independent Practice Association HMOs were included. We used Medicare data linked to tumor registry data from the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) program. Twelve cancer sites were studied: prostate; female breast; colon; bladder; rectum and rectosigmoid junction; corpus uteri and uterus, not otherwise specified; buccal cavity and pharynx; kidney and renal pelvis; melanoma of the skin; stomach; ovary; and cervix uteri. This study extends the work of earlier investigations by including cases from many HMOs, expanding the number of cancer sites studied, and including larger numbers of cases.

We selected these 12 cancer sites because of their varying degrees of potential for screening. Many HMOs with Medicare contracts cover services outside the regular Medicare benefit package, including preventive services, to induce beneficiaries to enroll. Preventive services, which often include cancer screenings and physical examinations, may lead to earlier stages at diagnosis for HMO enrollees.

Methods

Data

At the time of our study, the SEER program received uniformly reported data from nine tumor registries covering about 10% of the US population.[6] We included in our study the following geographic areas with significant Medicare HMO enrollments: Connecticut, Hawaii, Iowa, New Mexico, Detroit, San Francisco/Oakland, and Seattle. As population-based registries, SEER participants col-

Gerald F. Riley and James D. Lubitz are with the Health Care Financing Administration, Baltimore, Md. Arnold L. Potosky and Martin L. Brown are with the National Cancer Institute, Bethesda, Md.

Requests for reprints should be sent to Gerald F. Riley, MSPH, Health Care Financing Administration, Rm 2504, Oak Meadows Bldg, 6325 Security Blvd, Baltimore, MD 21207-5187.

This paper was accepted January 26, 1994.

Editor's Note. See related editorial by Robertson (p 1566) in this issue.

lect information on each incident cancer case in their reporting areas. Reported information includes month and year of diagnosis, site of cancer, stage at diagnosis, date of death, and county and census tract of residence.

Although SEER data do not constitute a probability sample of the nation, they are the primary source of national information on cancer incidence and survival.[6,7] SEER areas are concentrated in western states, and they involve a lower proportion of Blacks and a higher proportion of persons of "other" races than the US average.

SEER files were matched to Medicare enrollment files on an individual basis with a 94% match rate, as described elsewhere.[8] Medicare enrollment files contain entitlement dates to Part A and Part B, zip code of residence, and months in which the beneficiary was enrolled under a Medicare HMO risk or cost contract. Enrollees in both risk and cost HMOs were included in the study; findings for risk and cost contract enrollees were similar, and the results were pooled for this article. As of December 1989, 76% of Medicare HMO enrollees in SEER areas were in staff or group model HMOs, and 24% were Independent Practice Association HMOs.

Sample Selection

We selected first primary cases diagnosed between 1985 and 1989 in individuals who were 65 years of age or older and who were entitled to Part A and Part B at the time of diagnosis. For ovarian cancer, only cases diagnosed in 1985 through 1987 were selected because staging information was not available for later years at the time of our study. Lung cancer was excluded from the study because of the high number of unstaged cases.

We selected all cases diagnosed while the person was enrolled in an HMO. For each HMO case, two fee-for-service cases were randomly selected with the same cancer site, health service area of residence, and year of diagnosis (1985 through 1987 or 1988/89). A health service area is a geographic area, composed of one or more counties, that is relatively self-contained with respect to the provision of routine hospital care to Medicare beneficiaries.[9] Sample selection was based on health service area to control for geographic differences in provider practice patterns. In 1% of cases, there were insufficient fee-for-service cases in either a given health service area or year of diagnosis. In the former situation, fee-for-service cases were randomly selected from the appropriate registry rather than from the health service area; in the latter situation, fee-for-service cases were selected from other years of diagnosis. For melanomas, two fee-for-service cases were not available for each HMO case in San Francisco/Oakland, resulting in 76 (15%) too few fee-for-service cases from that registry.

TABLE 1—Number of Persons with Cancer and Percentage Distribution across Stage at Diagnosis, by Cancer Site and HMO Status: Aged Medicare Beneficiaries in SEER, 1985 to 1989

Cancer Site	All Persons	Distribution of Staged Cases, %				Unstaged, %
		In Situ	Local	Regional	Distant	
Strong evidence for screening						
Female breast[ab]						
HMO	2277	9.6	62.7	23.6	4.1	1.5
FFS	4507	8.6	57.4	28.4	5.7	3.9
Moderate evidence for screening						
Cervix uteri[ac]						
HMO	148	58.0	17.5	21.0	3.5	3.4
FFS	293	38.8	16.0	36.2	9.0	8.5
Colon[a]						
HMO	2039	5.9	41.5	36.6	16.0	3.6
FFS	4041	6.1	36.1	39.5	18.4	4.7
Rectum						
HMO	786	6.8	47.6	31.2	14.5	7.8
FFS	1567	7.5	43.5	34.4	14.7	6.9
Melanoma[ac]						
HMO	417	39.0	54.1	5.0	2.0	3.4
FFS	783	23.8	61.9	9.5	4.8	6.3
Weak evidence for screening						
Prostate[c]						
HMO	2737	0.0	64.3	16.8	18.8	8.8
FFS	5409	0.2	63.2	17.8	18.8	10.3
Buccal cavity and pharynx						
HMO	335	3.2	49.0	42.3	5.5	6.9
FFS	658	2.8	40.6	49.0	7.7	6.7
No evidence for screening						
Bladder						
HMO	880	. . .	76.6	20.0	3.5	4.9
FFS	1759	. . .	75.7	20.8	3.5	4.4
Corpus uteri and uterus, not otherwise specified[c]						
HMO	492	1.2	74.5	12.5	11.8	2.0
FFS	976	1.4	70.6	15.9	12.1	4.1
Kidney						
HMO	341	2.2	43.7	26.0	28.2	7.3
FFS	672	1.5	45.6	26.4	26.6	7.6
Stomach						
HMO	461	0.5	16.8	37.6	45.1	16.3
FFS	912	1.0	21.5	38.8	38.7	14.9
Ovary[d]						
HMO	146	0.0	11.0	34.3	54.7	6.2
FFS	278	0.4	11.9	37.7	50.0	6.5

Note. HMO = health maintenance organization; SEER = Surveillance, Epidemiology, and End Results program; FFS = fee-for-service.
[a]$P < .001$ for chi-square test of association between stage and HMO status.
[b]$P < .001$ for z test of difference in proportion unstaged between HMO and FFS.
[c]$P < .05$ for z test of difference in proportion unstaged between HMO and FFS.
[d]To compute chi-square values, the single in situ case was combined with local cases.

TABLE 2—Means and Percentage Distributions of Selected Independent Variables, by Patient's HMO Status: Aged Medicare Beneficiaries in SEER Diagnosed with Female Breast, Prostate, and Colon Cancer, 1985 to 1989

	Female Breast		Prostate		Colon	
	HMO	FFS	HMO	FFS	HMO	FFS
Mean age, y	73.2	74.4	74.3	75.0	75.0	76.2
Male (1 = yes)	0.00	0.00	1.00	1.00	0.53	0.48
Race, %						
White	84	83	77	77	78	77
Black	4	5	7	8	4	6
Asian (Chinese/Japanese/Filipino)	7	7	9	9	13	12
Other and unknown (includes Hispanic)	5	5	7	6	6	5
Married, %	46	40	75	71	59	52
Registry, %						
Connecticut	6	6	5	5	6	6
San Francisco/Oakland	49	49	44	43	46	45
Detroit	5	5	6	6	6	6
Hawaii	11	11	13	13	15	15
Iowa	6	6	7	7	6	6
New Mexico	4	4	6	7	5	5
Seattle	18	18	19	20	16	16
Adults with < 12 years of education (census tract/zip code), %	23	23	24	24	24	25

Note. Unstaged cases are excluded. HMO = health maintenance organization; SEER = Surveillance, Epidemiology, and End Results program; FFS = fee-for-service.

Staging

The NCI codes stage at diagnosis using extent of disease information reported by the registries. The registries abstract information from a variety of sources, including inpatient hospital records, outpatient records, and pathology reports. The staging system recognizes four stages: in situ indicates a noninvasive malignancy; local stage indicates an invasive cancer confined to the site/organ of origin; regional stage indicates spread by direct extension to adjacent organs or to regional lymph nodes; and distant stage indicates spread to distant organs or lymph nodes.

Between 3% and 15% of cases were unstaged, depending on cancer site, because there was inadequate information available on extent of disease. Unstaged cases involve relatively high death rates, suggesting that they consist largely of regional and distant cases.[6]

Analysis

We first excluded cases diagnosed through death certificate review or autopsy (1.2% of all HMO cases and 2.1% of fee-for-service cases) and all unstaged cases. We then estimated three separate logistic regression equations for each cancer site using the following dependent variables: Y1 (1 if stage is distant, 0 for earlier stages), Y2 (1 if stage is regional, 0 if stage is in situ or local), and Y3 (1 if stage is local, 0 if stage is in situ).

Distant stage cases were excluded from the second model, and both distant and regional stage cases were excluded from the third. This method has been described by Fienberg[10] and is appropriate in situations in which the levels of the dependent variable are ordered. If there were fewer than 60 distant stage cases for a given site, those cases were pooled with the regional cases. Unconditional logistic regression was used because most site, health service area, and year strata were large, making the use of conditional logistic regression unnecessary.[11]

Odds ratios (ORs) and 95% confidence intervals were estimated according to standard methods from logistic regression modeling.[12] Goodness of fit of the regression models was assessed by comparing observed and expected numbers of cases with late and early stage disease within deciles of estimated probabilities.[13]

The equations incorporated the following independent variables: age, sex, race, marital status, percentage of adults not completing high school at the census tract level, health service area, year of diagnosis, and HMO enrollment status. Age, race, and marital status were included because previous studies have shown an association between these variables and stage for some cancer sites.[14–18] Race was coded as White, Black, Asian (Chinese/Japanese/Filipino), or other.

Previous studies have shown that lower socioeconomic status (SES) is associated with later stage at diagnosis for breast and cervical cancers.[19–21] Education at the census tract level (expressed as the proportion of adults with less than 12 years of education), based on US census data, was used as a proxy for individual SES status. Education is related conceptually and empirically to both knowledge and use of cancer screening tests.[22] In untracted areas, education at the zip code level was used (14% of cases).

Classification of Cancer Sites

We classified cancer sites on the basis of the efficacy of screening for reducing mortality. Evidence on the efficacy of cancer screening is summarized by the Physician Data Query system, a computerized clinical information service on cancer maintained by NCI.[23] Five levels of evidence are cited, ranging from properly randomized, well-designed and conducted, controlled trials (level 1) to opinions of respected authorities based on clinical experience, descriptive studies, or reports of expert committees (level 5). On the basis of Physician Data Query levels of evidence at the time of our study, we classified cancer of the female breast as having strong evidence for screening; cancers of the cervix uteri, colon, and rectum and melanomas as having moderate evidence; cancers of the prostate and buccal cavity and pharynx as having weak evidence; and cancers of the bladder, uterus, kidney, stomach, and ovary as having no evidence.

Results

HMO enrollees were diagnosed at significantly earlier stages for four sites without controlling for covariates: female breast, cervix, colon, and melanoma (Table 1). The largest difference between HMO enrollees and nonenrollees (i.e., beneficiaries in fee-for-service) was for women with cervical cancer. Fifty-eight percent of HMO enrollees with cervical cancer were

diagnosed at the in situ stage, in comparison with only 38.8% of nonenrollees. There were also large differences associated with melanomas, with 39% of HMO enrollees diagnosed at the in situ stage and only 23.8% of nonenrollees diagnosed at that stage. Differences in stage between enrollees and nonenrollees were smaller in magnitude for breast and colon cancer but were statistically significant in part because of their larger sample sizes.

The percentage of cases for which stage was known varied among sites (Table 1). For most sites, fewer than 10% of cases were unstaged; for stomach and prostate cancers, however, approximately 15% and 10% respectively, were unstaged. There was a significantly smaller percentage of unstaged cases among HMO enrollees than among nonenrollees for five sites: female breast, cervix, prostate, melanoma, and uterus.

Table 2 summarizes characteristics of HMO enrollees and nonenrollees with female breast, prostate, and colon cancers. There were few differences. The average age of HMO enrollees tended to be about 1 year less than that of nonenrollees. There were no important differences by race, registry, or proportion of adults with less than 12 years of education at the census tract or zip code level. More HMO enrollees tended to be married than nonenrollees, reflecting in part their younger age.

About 45% of sample persons were in the San Francisco/Oakland area, reflecting the high enrollment in Kaiser Permanente. Nearly 20% resided in the Seattle area, most of whom were enrolled in the Group Health Cooperative of Puget Sound. The Hawaii registry had the third highest number of sample persons, resulting in a geographic bias toward HMOs in the western United States.

HMO enrollees with breast cancer were less likely to be diagnosed at distant stages than were nonenrollees (OR = 0.73) (Table 3). HMO enrollees were also less likely to be diagnosed at regional stages than at earlier stages (OR = 0.78). Among women diagnosed at the in situ or local stage, there was no effect of HMO enrollment on stage.

HMO enrollees with cervical cancer were only about one third as likely as nonenrollees to be diagnosed at the regional or distant stage (OR = 0.35). The odds ratio for local vs in situ disease was also below one but was not significant.

Among persons with colon cancer, HMO enrollees were less likely to be diagnosed at regional stages than were nonenrollees (OR = 0.85). HMO enrollees with melanoma were less than half as likely as nonenrollees to be diagnosed at the regional or distant stage (OR = 0.44); among cases diagnosed at the in situ or local stage, HMO enrollees were only about half as likely to be diagnosed at the local stage (OR = 0.52).

For persons with stomach cancer, HMO enrollment was associated with distant stage disease (OR = 1.32). The odds ratio for regional vs in situ or local disease was also above one but was not significant.

Chi-square statistics indicating goodness of fit were nonsignificant at the 5% level for all regression equations except one, indicating acceptable fit of the models to the data. (The full results of the logistic regression analyses are available from the first author.)

Because nearly half of the HMO enrollees were in the San Francisco/Oakland registry (most of whom were Kaiser Permanente enrollees), our findings may not represent a general HMO effect. To test the generalizability of our findings, we developed separate models for the following registries for breast and colon cancers: San Francisco/Oakland, Seattle, Hawaii, and all other registries combined. The HMO effects were strongest in the San Francisco/Oakland and Seattle areas (Table 4). For breast cancer cases in San Francisco/Oakland and Seattle, HMO enrollees were diagnosed at significantly earlier stages. For "other" registries, there was a nonsignificant effect of fewer regional stages for HMO enrollees. For Hawaii, there was no pattern of earlier stage for HMO enrollees.

For colon cancer, HMO enrollees were diagnosed at significantly earlier stages in San Francisco/Oakland, Seattle,

TABLE 3—Odds Ratios for the Association of HMO Enrollment Status with Cancer Stage at Diagnosis: Aged Medicare Beneficiaries in SEER Diagnosed from 1985 to 1989

Site of Cancer	Distant vs Earlier Stages: OR for Distant Stage	95% CI	Regional vs Earlier Stages: OR for Regional Stage	95% CI	Local vs In Situ Stage: OR for Local Stage	95% CI
Strong evidence for screening						
Female breast	0.73	0.57, 0.94	0.78	0.69, 0.87	1.01	0.84, 1.21
Moderate evidence for screening						
Cervix uteri	NE		0.34	0.21, 0.56	0.73	0.39, 1.35
Colon	0.87	0.75, 1.01	0.85	0.75, 0.96	1.20	0.94, 1.53
Rectum	0.98	0.76, 1.27	0.86	0.70, 1.06	1.25	0.86, 1.81
Melanoma	NE		0.44	0.28, 0.68	0.52	0.39, 0.69
Weak evidence for screening						
Prostate	1.04	0.92, 1.18	0.91	0.79, 1.04	NE	
Buccal cavity and pharynx	0.79	0.43, 1.43	0.75	0.56, 1.02	NE	
No evidence for screening						
Bladder	1.06	0.67, 1.69	0.99	0.80, 1.22	. . .	. . .
Corpus uteri and uterus, not otherwise specified	1.03	0.73, 1.46	0.79	0.57, 1.11	NE	
Kidney	1.12	0.82, 1.53	1.02	0.73, 1.44	NE	
Stomach	1.32	1.03, 1.70	1.18	0.83, 1.69	NE	
Ovary	1.32	0.84, 2.06	NE		NE	

Note. For cervical cancer and melanoma, distant and regional cases were pooled because of the small number of distant stage cases. HMO = health maintenance organization; SEER = Surveillance, Epidemiology, and End Results program; NE = not estimated; OR = odds ratio; CI = confidence interval.

TABLE 4—Odds Ratios for the Association of HMO Enrollment Status with Stage at Diagnosis: Aged Medicare Beneficiaries in SEER Diagnosed with Female Breast and Colon Cancer, 1985 to 1989

Geographic Area	Distant vs. Earlier Stages: OR for Distant Stage	95% CI	Regional vs. Earlier Stages: OR for Regional Stage	95% CI	Local vs. In Situ Stage: OR for Local Stage	95% CI
Female breast						
San Francisco/ Oakland	0.59	0.39, 0.87	0.74	0.63, 0.88	0.92	0.71, 1.19
Seattle	NE		0.66	0.50, 0.87	1.27	0.82, 1.97
Hawaii	NE		1.09	0.77, 1.55	1.07	0.65, 1.76
Other[a]	0.92	0.59, 1.44	0.81	0.63, 1.04	1.12	0.72, 1.73
Total	0.73	0.57, 0.94	0.78	0.69, 0.87	1.01	0.84, 1.21
Colon						
San Francisco/ Oakland	0.92	0.74, 1.14	0.80	0.67, 0.95	1.56	1.05, 2.33
Seattle	0.63	0.43, 0.92	0.80	0.59, 1.08	1.48	0.82, 2.65
Hawaii	0.59	0.39, 0.91	0.90	0.65, 1.24	1.03	0.60, 1.76
Other[a]	1.18	0.88, 1.57	0.99	0.77, 1.28	0.77	0.46, 1.29
Total	0.87	0.75, 1.01	0.85	0.75, 0.96	1.20	0.94, 1.53

Note. For breast cancer cases in Seattle and Hawaii, distant and regional cases were pooled. HMO = health maintenance organization; SEER = Surveillance, Epidemiology, and End Results program; NE = not estimated separately because of small sample size; OR = odds ratio; CI = confidence interval.

[a]Connecticut, Detroit, Iowa, New Mexico.

and Hawaii but not in the other registries combined.

Discussion

Screening Services and Stage at Diagnosis

HMO enrollees were diagnosed at earlier stages for four sites—female breast, cervix, colon, and melanomas—for which there is strong or moderate evidence for screening. The earlier stages we found may be attributable to HMO coverage of procedures such as screening mammograms, clinical breast examinations, Pap smears, fecal occult blood tests, sigmoidoscopies, and physical examinations. (Medicare did not cover Pap smears until July 1990 and screening mammograms until January 1991.) More than 90% of HMOs with Medicare risk contracts include coverage of preventive services in their benefit packages.[1] Kaiser Permanente and the Group Health Cooperative of Puget Sound, the two largest sources of HMO enrollees in our study, provide preventive services for their Medicare enrollees. Most preventive services are not covered under Medicare fee-for-service, and such services are seldom covered under supplemental "Medigap" policies.[24] Studies of service use have found that HMO enrollees receive more cancer screening services than nonenrollees.[22,25,26] The greater availability of screening services in HMOs may be particularly important for the elderly because elderly women use screening mammographies and Pap smears less frequently than do younger women.[27]

Medicare coverage of Pap smears and screening mammograms under fee-for-service began in 1990 and 1991, respectively. If use of these services is increasing because of Medicare coverage, elderly women may experience earlier stages at diagnosis over time for cancers of the cervix and breast. Future analyses of SEER data can reveal whether this is occurring and whether the distribution of stage at diagnosis in fee-for-service becomes more like that in HMOs over time. Burg and Lane[28] argue that Medicare payment for screening mammography is unlikely to substantially increase use, however, based on physician attitudes toward mammography for older women.

We found that HMO enrollees were diagnosed at significantly earlier stages for melanomas. This type of cancer may be amenable to detection during a routine physical examination or other physician encounter. Although Medicare does not cover routine physical examinations under fee-for-service, 82% of HMOs with risk contracts covered such examinations for their Medicare enrollees as of December 1989. Brown et al.[1] found that HMO enrollees were somewhat more likely than nonenrollees to have a physical examination and were more likely to have at least one physician visit in a year.

Although previous studies have shown that HMO enrollees tend to use more preventive services than nonenrollees, it is not clear whether this is due to insurance coverage of such services or to promotion of their use by HMOs. Luft,[29] in a review of the literature, suggested that differential use of preventive services is due to an "insurance effect" (i.e., financial coverage) and not to an HMO "health maintenance effect." Manning et al.,[30] however, found evidence for both effects. Brown et al.[22] found that HMO membership was associated with both knowledge and use of fecal occult blood tests, suggesting that part of the HMO effect on use of screening services may be due to preventive health education, as well as increased access. Another explanation, however, is that prevention-oriented individuals may have self-selected into HMOs. Such prevention-oriented persons may have sought out screening services even if they had not enrolled.

With the exception of stomach cancer, we did not find differences in stage between HMOs and fee-for-service for sites lacking routine screening procedures. This finding is consistent with a lack of systematic differences between HMO and fee-for-service care in access to services related to cancer diagnosis for those sites lacking routine screening procedures. It is also consistent with a lack of systematic differences between enrollees and nonenrollees in their care-seeking behavior for symptoms indicative of cancer.

Other studies have tended to find no differences in stage at diagnosis between HMOs and fee-for-service. A reason that our findings differ from those of other studies may be that our study was restricted to the elderly population, most of whom did not have insurance coverage of preventive services outside an HMO setting. We also had much larger sample sizes than were available in previous studies.

Although we found earlier stages at diagnosis for HMO enrollees, we did not determine whether survival was improved for them. Screening tests may produce a shift to earlier stages at diagnosis without

any improvement in survival because of length and lead time bias effects.[31] Randomized controlled trials of screening mammography and fecal occult blood tests have demonstrated reductions in mortality accompanied by shifts in stage at diagnosis.[32,33] An analysis of survival experience was beyond the scope of our study.

Limitations

Stage at diagnosis is determined from information on extent of disease, which may be influenced by the diagnostic and therapeutic procedures used, as well as by the adequacy of provider record keeping. Any differences in such factors between HMOs and fee-for-service providers could influence our findings. The higher percentage of unstaged cases in fee-for-service suggests that reporting of extent of disease information was better for HMO enrollees. The high mortality of unstaged cases suggests possible underidentification of regional and distant cases, which would make our findings conservative regarding earlier diagnosis in HMOs. It is also possible, however, that the high mortality of unstaged cases may be due to the poorer underlying health involved in these cases rather than to greater spread of cancer at the time of diagnosis.

Our findings may not be generalizable to all HMOs with Medicare contracts. The HMO enrollees in our study were disproportionately enrolled in staff and group model HMOs. In addition, most HMO enrollees in San Francisco/Oakland and many enrollees in Hawaii were members of Kaiser Permanente; most in Seattle were members of the Group Health Cooperative of Puget Sound. Our database did not permit us to identify in which HMOs specific individuals were enrolled.

Policy Implications

Our findings suggest that coverage of preventive services may affect stage of cancer at diagnosis, which is strongly associated with survival and morbidity. Policymakers should consider a significant role for preventive services under health care reform. The Health Security Act, for example, proposes coverage for a significant level of preventive care, including Pap smears, screening mammograms, and clinician visits. Such services can be provided at reasonable cost, as evidenced by many HMOs that provide such services to Medicare enrollees and remain competitive with fee-for-service.

Stage at diagnosis can serve as a measure of access and quality of care in individual health plans. Under the Health Security Act, health alliances are expected to have a strong role in evaluating the quality of care in plans and making their findings available to consumers. Monitoring stage of cancer at diagnosis may be part of the alliances' evaluation function, particularly if Medicare comes under their purview. Monitoring stage may be somewhat easier than monitoring outcomes of care, which usually require complex case mix adjustments. Finally, health alliances should directly monitor access to and use of preventive services among participating health plans because access to such services is an important dimension of plan quality. □

Acknowledgments

We wish to thank Barry Bye for his assistance with statistical aspects of the study. We also appreciate the helpful comments of Carol Kosary and Karen Johnson, who reviewed earlier drafts of the paper. We thank Bruce Fireman of Kaiser Permanente and Stephen Taplin of the Group Health Cooperative of Puget Sound for providing information on the preventive services covered by their respective plans. We also thank Jennifer Stevens for her programming support and Florence Beckman and Charlotte Smith for their help in manuscript preparation.

References

1. Brown R, Bergeron J, Clement DG, Hill JW, Retchin SM. *The Medicare Risk Program for HMOs—Final Summary Report on Findings from the Evaluation.* Princeton, NJ: Mathematica Policy Research; 1993. Final report to HCFA under contract 500-88-0006.
2. Kulkarni PR, Vernon SW, Jackson GL, Lairson D, Davis BR. Stage at diagnosis of breast cancer: comparison in a fee-for-service and health maintenance organization practice. *Med Care.* 1989;27:608–621.
3. Greenwald HP. HMO membership, copayment, and initiation of care for cancer: a study of working adults. *Am J Public Health.* 1987;77:461–466.
4. Greenwald HP, Henke CJ. HMO membership, treatment, and mortality risk among prostatic cancer patients. *Am J Public Health.* 1992;82:1099–1104.
5. Francis AM, Polissar L, Lorenz AB. Care of patients with colorectal cancer: a comparison of a health maintenance organization and fee-for-service practices. *Med Care.* 1984;22:418–429.
6. Miller BA, Ries LAG, Hankey BF, et al., eds. *Cancer Statistics Review: 1973–1990.* Bethesda, Md: National Cancer Institute; 1993. NIH publication 93-2789.
7. Boring CC, Squires TS, Tong T. Cancer statistics, 1993. *CA.* 1993;43:7–26.
8. Potosky AL, Riley G, Lubitz J, Mentnech R, Kessler L. Potential for cancer related health services research using a linked Medicare-tumor registry data base. *Med Care.* 1993;31:732–748.
9. Makuc DM, Haglund B, Ingram DD, Kleinman JC, Feldman JJ. Health service areas for the U.S. *Vital Health Stat [2].* 1991;112. PHS publication 91-1386.
10. Fienberg SE. *The Analysis of Cross-Classified Categorical Data.* Cambridge, Mass: The MIT Press; 1992:86–90.
11. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods.* New York, NY: Van Nostrand Reinhold; 1982:439–440.
12. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons Inc; 1989:39–81.
13. Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol.* 1982;115:92–106.
14. Goodwin JS, Samet JM, Key CR, Humble C, Kutvirt D, Hunt C. Stage at diagnosis of cancer varies with the age of the patient. *J Am Geriatr Soc.* 1986;34:20–26.
15. Diehr P, Yergan J, Chu J, et al. Treatment modality and quality differences for black and white breast-cancer patients treated in community hospitals. *Med Care.* 1989;27:942–954.
16. McWhorter WP, Mayer WJ. Black/white differences in type of initial breast cancer treatment and implications for survival. *Am J Public Health.* 1987;77:1515–1517.
17. Mebane C, Gibbs T, Horm J. Current status of prostate cancer in North American black males. *J Natl Med Assoc.* 1990;82:782–788.
18. Goodwin JS, Hunt WC, Key CR, Samet JM. The effect of marital status on stage, treatment, and survival of cancer patients. *JAMA.* 1987;258:3125–3130.
19. Farley TA, Flannery JT. Late-stage diagnosis of breast cancer in women of lower socioeconomic status: public health implications. *Am J Public Health.* 1989;79:1508–1512.
20. Wells BL, Horm JW. Stage at diagnosis in breast cancer: race and socioeconomic factors. *Am J Public Health.* 1992;82:1383–1385.
21. Mandelblatt J, Andrews H, Kerner J, Zauber A, Burnett W. Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type. *Am J Public Health.* 1991;81:646–649.
22. Brown ML, Potosky AL, Thompson GB, Kessler LG. The knowledge and use of screening tests for colorectal and prostate cancer: data from the 1987 National Health Interview Survey. *Prev Med.* 1990;19:562–574.
23. *Cancer Screening Guidelines.* Bethesda, Md: National Cancer Institute; 1993.
24. Rice T, Thomas K. Evaluating the new Medigap standardization regulations. *Health Aff.* Spring 1992:194–207.
25. Bernstein AB, Thompson GB, Harlan LC. Differences in rates of cancer screening by usual source of medical care. *Med Care.* 1991;29:196–209.
26. Zapka JG, Hosmer D, Costanza ME, Harris DR, Stoddard A. Changes in mammography use: economic, need, and service factors. *Am J Public Health.* 1992;82:1345–1351.
27. Calle EE, Flanders WD, Thun MJ, Martin LM. Demographic predictors of mammog-

raphy and Pap smear screening in US women. *Am J Public Health.* 1993;83:53–60.
28. Burg MA, Lane DS. Mammography referrals for elderly women: is Medicare reimbursement likely to make a difference? *Health Services Res.* 1992;27:505–515.
29. Luft HS. Assessing the evidence on HMO performance. *Milbank Q.* 1980;58:501–536.
30. Manning WG, Leibowitz A, Goldberg GA, Rogers WH, Newhouse JP. A controlled trial of the effect of a prepaid group practice on use of services. *N Engl J Med.* 1984;310:1505–1510.
31. Prorok PC, Hankey BF, Bundy BN. Concepts and problems in the evaluation of screening programs. *J Chronic Dis.* 1981;34:159–171.
32. Chu KC, Smart CR, Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan Study: a randomized trial with breast cancer screening. *JNCI.* 1988;80:1125–1132.
33. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med.* 1993;328:1365–1371.

Report Available on the Rapid Identification and Treatment of Acute Myocardial Infarction

Each year, about 1.25 million people in the United States experience a heart attack. Approximately 500 000 of these people die. It is thought that many of these deaths could be prevented by initiating medical help as soon as a heart attack is suspected. A new report by the National Heart Attack Alert Program (NHAAP), *Emergency Department: Rapid Identification and Treatment of Patients with Acute Myocardial Infarction,* discusses the scientific basis for early care and makes a number of recommendations for reducing emergency department delays in the identification and treatment of heart attack patients.

The report concentrates on the emergency department's contribution to overall time until treatment for acute myocardial infarction (AMI). It describes the scientific basis for decreasing the time from the onset of symptoms to treatment and then specifically addresses impediments to rapid care in the emergency department. The report proposes changes in processes that can reduce delays at each point and presents an AMI time-to-treatment flow sheet for tracking these critical AMI timepoints. The report encourages all departments responsible for acute care to coordinate and establish protocols to ensure rapid identification and treatment of AMI patients. Finally, the report challenges emergency departments to implement a system of continuous quality improvement to identify and reduce delays in the treatment of AMI patients.

The NHAAP was launched by the National Heart, Lung, and Blood Institute in June 1991 to promote the rapid identification and treatment of individuals with symptoms and signs of AMI. The overall goal of the program is to reduce AMI-related morbidity and mortality. The NHAAP Coordinating Committee—consisting of representatives from 40 national professional, scientific, governmental, and voluntary organizations (including APHA)—was formed to help guide the program.

For a complimentary copy of the report (NIH publication 93-3278), contact the National Heart, Lung, and Blood Institute Information Center, PO Box 30105, Bethesda, MD 20824-0105; tel (301) 251-1222; fax (301) 251-1223.

Acutely Injured Patients with Trauma in Massachusetts: Differences in Care and Mortality, by Insurance Status

ABSTRACT

Objectives. This study was designed to determine whether resource use and mortality differed by insurance status for patients with acute trauma.

Methods. All adults emergently hospitalized in Massachusetts during 1990 with acute trauma (n = 15 008) were examined.

Results. After adjustment for confounders, uninsured patients were as likely to receive care in an intensive care unit as were patients with private insurance (odds ratio [OR] = 0.97, 95% confidence interval [CI] = 0.85, 1.11) but were less likely to undergo an operative procedure (OR = 0.68, 95% CI = 0.63, 0.74) or physical therapy (OR = 0.61, 95% CI = 0.57, 0.67) and were more likely to die in a hospital (OR = 2.15, 95% CI = 1.44, 3.19). Compared with patients with private insurance, those with Medicaid were less likely to receive an operative procedure (0.85, 0.75–0.97), were equally likely to receive care in an intensive care unit (OR = 1.05, 95% CI = 0.86, 1.30) or physical therapy (OR = 0.90, 95% CI = 0.79, 1.02), and were no more likely to die (OR = 1.28, 95% CI = 0.69, 2.39).

Conclusions. These results suggest that the uninsured receive less trauma-related care and have a higher mortality rate. The excess mortality in uninsured patients may be avoided if their resource use is increased to that of insured patients. (*Am J Public Health.* 1994;84:1605–1608)

Jennifer S. Haas, MD, MSPH, and Lee Goldman, MD, MPH

Introduction

Trauma is the fourth leading cause of death in the United States,[1] and it results in substantial morbidity. The annual cost of traumatic injuries is estimated to be more than $100 billion.[2] Some data suggest that 30% of trauma-related deaths may be prevented by appropriate prehospital and in-hospital care.[3,4]

Although previous work has demonstrated that the uninsured[5–8] are less likely to receive certain health care services, these studies have usually examined elective procedures or ambulatory services for which differences in rates of use may reflect patients' preferences. In addition, it has been difficult to control for differences in disease severity for chronic medical problems.[9] Finally, differences in procedure rates may not be associated with differences in health outcomes. Because of these limitations, no conclusions can be drawn as to the appropriateness of the observed variation.

To address whether lower rates of resource use by uninsured patients are associated with worse health outcomes, we studied patients with acute trauma. Since these patients are acutely injured, the relationship between in-hospital care and mortality may be more direct than for chronic health problems. Also, patients' preferences are less likely to explain differences in the use of procedures. Furthermore, the severity of injury can be measured reliably for patients with traumatic injuries in administrative data sets. The goal of this study was to examine whether resource use and mortality differed by insurance status for patients hospitalized with acute trauma.

Methods

Data

Discharge abstracts for all Massachusetts residents 15 to 64 years of age emergently hospitalized in a nonfederal, acute care hospital in Massachusetts during fiscal year 1990 were obtained from the Massachusetts Health Data Consortium. *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM)[10] codes were used to indicate diagnoses and procedures. A maximum of 10 diagnosis and 10 procedure codes were listed for each patient. Of the 94 acute care hospitals in Massachusetts with trauma discharges in 1990, 9 were designated as level 1 trauma centers. Hospitals located in Boston, Cambridge, Springfield, Worcester, Lowell, or New Bedford were considered to be in an urban area. All of the level 1 trauma centers were located in an urban area.

Patients with a principal diagnosis code indicating major traumatic injury (ICD-9-CM codes 800 to 959), excluding those whose codes indicated the late effects of trauma (codes 905 to 909) or a foreign body (codes 930 to 939), were included in this study. Abstracts that indicated a transfer from another hospital were excluded to avoid analyzing multiple abstracts for the same patient.

Variables

Insurance status. Patients were categorized into three groups: (1) uninsured (self-pay or free care), (2) Medicaid, and (3) private insurance (Blue Cross/Blue Shield, commercial payers, and prepaid health plans).

The authors are with the Division of General Medicine, Section for Clinical Epidemiology, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Mass.

Requests for reprints should be sent to Jennifer S. Haas, MD, MSPH, Division of General Medicine, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115.

This paper was accepted May 10, 1994.

Editor's Note. See related editorial by Robertson (p 1566) in this issue.

TABLE 1—Characteristics of the Study Population (n = 15 008)

	Sample
Mean age, y (range)	33.7 (15–64)
Race, no. (%)	
White	12 102 (81)
Non-White	2 906 (19)
Payer, no. (%)	
Uninsured	4 735 (32)
Medicaid	1 254 (8)
Private	9 019 (60)
Injury Severity Score, mean (range)	12.4 (2–76)
At least one comorbid condition, no. (%)	639 (4%)
Process of care, no. (%)	
Operative procedure	7 120 (47%)
Intensive care unit	1 757 (12%)
Physical therapy	5 902 (39%)
In-hospital mortality, no. (%)	176 (1.2%)

TABLE 2—Adjusted Odds Ratios (95% Confidence Intervals) for Admission to a Level 1 Trauma Center or a Hospital Located in an Urban Area, by Payer

Payer	Admission to a Level 1 Trauma Center	Admission to a Hospital in an Urban Area
Uninsured	1.51 (1.39, 1.64)	1.45 (1.34, 1.57)
Medicaid	1.32 (1.15, 1.51)	1.35 (1.18, 1.54)
Private	1.0 ...	1.0 ...

Note. Odds ratios were adjusted for age, sex, race, Injury Severity Score, and comorbidity.

Resource use and in-hospital mortality. Measures of resource use that can be reliably identified in discharge abstracts include whether a patient has had an operative procedure, has been cared for in an intensive care unit, or has received physical therapy. These measures of resource use also have "face validity" (i.e., each is generally believed to improve health outcomes). A patient was considered to have undergone an operative procedure or physical therapy if the corresponding charge was noted or to have had an intensive care unit stay if a stay of at least 1 day in a medical, surgical, or cardiac intensive care unit was indicated. In-hospital mortality was defined as a discharge disposition of expired.

Race, severity of injury, comorbidity, and mechanism of injury. Patients were characterized into two racial groups, White and non-White (African American, Latino, Asian, other). The Abbreviated Injury Scale,[11] which assesses the severity of injury of each body region and is the most widely accepted injury severity measure based on anatomic injury,[12] was used to measure severity of injury. ICD-9-CM diagnosis codes were converted into Abbreviated Injury Scale scores by means of a validated algorithm.[13] The Injury Severity Score, which summarizes the Abbreviated Injury Scale scores of the three most severely injured body regions and has been associated with mortality from trauma,[14–16] was used to categorize overall severity of injury.

Conditions that have previously been shown to affect mortality in patients with trauma were analyzed,[17] including secondary diagnosis codes for cirrhosis (ICD-9-CM code 571), congenital coagulopathy (codes 286 and 287, excluding 286.6 and 287.4), ischemic heart disease (codes 410 to 414), chronic obstructive pulmonary disease (codes 490 to 496), and diabetes (code 250). Patients with at least one of these codes were defined as having a comorbid condition.

Information on mechanism of injury was available for 4873 patients (32.5%). Mechanism of injury was categorized as a gunshot wound (ICD-9-CM code E922, E955, E965, E970, or E985), a stab wound (code E920, E956, E966, E974, or E986), a fall (codes E880 to E888), a burn (codes 940 to 949), or a motor vehicle accident (codes E810 to E829). Patients with a diagnosis code indicating a gunshot wound or a stab wound were defined as having a penetrating injury.

Data Analysis

Differences in severity of injury (Injury Severity Score) between payer groups were analyzed with analysis of variance. Logistic regression analyses[18] were used to examine the relationship between payer and each marker of trauma-related care and mortality while adjusting for possible confounders, including age, sex, race, Injury Severity Score, and comorbidity. The same approach was used, in subgroup analyses, to examine patients with severe injuries (Injury Severity Score greater than 12), patients admitted to a hospital located in an urban area, and patients admitted to a level 1 trauma center. For the subgroup of patients with information about mechanism of injury, models that also controlled for penetrating vs nonpenetrating injury were examined.

Results

Most patients hospitalized with acute trauma in Massachusetts were White and privately insured (Table 1). The mean Injury Severity Score was 12.4. For example, the most common principal diagnosis associated with an Injury Severity Score of 12 was a concussion associated with a loss of consciousness. Of the 7120 patients (47%) who underwent an operative procedure, 5719 (81%) did so within the first 24 hours of hospitalization. The mean Injury Severity Scores were 12.4 for patients with private insurance, 12.8 for those with Medicaid, and 12.3 for the uninsured ($P = .20$).

The Injury Severity Score was significantly associated with resource use and in-hospital mortality. The mean Injury Severity Score for patients who underwent an operative procedure was 13.9, in comparison with 11.0 for those who did not ($P < .001$). For patients with an intensive care unit stay, the mean score was 22.4, in comparison with 11.0 for patients without such a stay ($P < .001$); for patients who received physical therapy, the mean score was 13.8, in comparison with 11.4 for those who did not ($P < .001$). The mean injury severity score for patients who died in the hospital was 41.8 in comparison with 12.0 for those patients who survived ($P < .001$).

Overall, 7111 patients (47%) received care in a hospital located in an urban area, and 4779 (32%) received care in a level 1 trauma center. Patients who were uninsured or covered by Medicaid were more likely to receive care in an urban hospital or in a level 1 trauma center after adjustment for age, race, injury severity score, and comorbidity (Table 2).

For the subgroup of patients with a known mechanism of injury, the Injury Severity Score for patients with a penetrating injury was not quite significantly

different than that for patients with other types of trauma (13.2 and 12.5, respectively, $P = .06$). The uninsured were more likely than patients with private insurance to have a penetrating injury (32.5% vs 10.6%, $P < .001$). After adjustment for Injury Severity Score, patients with a penetrating injury were more likely to die in the hospital (odds ratio [OR] = 2.46, 95% confidence interval [CI] = 1.36, 4.43).

After adjustment for age, sex, race, severity of injury, and comorbidity, patients who were uninsured, on average, received fewer resources and had higher rates of mortality than patients with private insurance (Table 3). Although patients with Medicaid were less likely to undergo an operative procedure than patients with private insurance, they were just as likely to receive care in an intensive care unit or to receive physical therapy, and they were not significantly more likely to die in the hospital.

Since patients who die early in their hospital stay may not have a full "opportunity" to receive trauma-related care, the analyses were repeated after excluding patients who died within the first 48 hours of hospitalization. After these patients were excluded, resource use was similar in magnitude and significance to that presented in Table 3.

In general, the patterns of trauma-related care and mortality were similar in subgroups of patients who were admitted to level 1 trauma centers, who were admitted to a hospital located in an urban area, who had severe injuries, or who had a known mechanism of injury as in the overall population (Table 4).

TABLE 3—Adjusted Odds Ratios (95% Confidence Intervals) for Trauma-Related Care and In-Hospital Mortality, by Payer

Payer	Use of an Intensive Care Unit	Operative Procedure	Physical Therapy	In-Hospital Mortality
Uninsured	0.97 (0.85, 1.11)	0.68 (0.63, 0.74)	0.61 (0.57, 0.67)	2.15 (1.44, 3.19)
Medicaid	1.05 (0.86, 1.30)	0.85 (0.75, 0.97)	0.90 (0.79, 1.02)	1.28 (0.69, 2.39)
Private	1.0 ...	1.0 ...	1.0 ...	1.0 ...

Note. Odds ratios were adjusted for age, sex, race, Injury Severity Score, and comorbidity.

TABLE 4—Adjusted Odds Ratios (95% Confidence Intervals) for Trauma-Related Care and In-Hospital Mortality, by Payer, for Selected Patient Subgroups

Payer	Use of an Intensive Care Unit	Operative Procedure	Physical Therapy	In-Hospital Mortality
Severe injury[a] (n = 4102)				
Uninsured	0.91 (0.76, 1.09)	0.78 (0.67–0.91)	0.61 (0.53–0.71)	1.90 (1.22–2.97)
Medicaid	1.08 (0.83–1.42)	1.15 (0.92–1.44)	1.39 (1.11–1.75)	1.08 (0.53–2.18)
Private	1.0 ...	1.0 ...	1.0 ...	1.0 ...
Known mechanism of injury (n = 4873)[b]				
Uninsured	0.88 (0.69–1.13)	0.81 (0.71–0.93)	0.77 (0.66–0.90)	2.01 (1.02–3.97)
Medicaid	1.09 (0.74–1.61)	0.84 (0.67–1.04)	0.98 (0.78–1.24)	1.12 (0.40–3.16)
Private	1.0 ...	1.0 ...	1.0 ...	1.0 ...
Admitted to a level 1 trauma center (n = 4779)				
Uninsured	0.71 (0.55–0.91)	0.70 (0.61–0.80)	0.51 (0.44–0.59)	2.46 (1.47–4.13)
Medicaid	1.05 (0.73–1.52)	0.79 (0.64–0.97)	0.78 (0.64–0.97)	1.52 (0.71–3.23)
Private	1.0 ...	1.0 ...	1.0 ...	1.0 ...
Admitted to a hospital located in an urban area (n = 7111)				
Uninsured	0.84 (0.69–1.03)	0.62 (0.56–0.69)	0.53 (0.48–0.60)	1.94 (1.22–3.07)
Medicaid	1.20 (0.90–1.61)	0.78 (0.66–0.93)	0.80 (0.68–0.96)	1.23 (0.61–2.48)
Private	1.0 ...	1.0 ...	1.0 ...	1.0 ...

Note. Odds ratios were adjusted for age, sex, race, Injury Severity Score, and comorbidity.
[a]Injury Severity Score greater than 12.
[b]Models for patients with a known mechanism of injury were adjusted for age, sex, race, Injury Severity Score, comorbidity, and mechanism of injury (penetrating vs nonpenetrating injury).

Discussion

This study suggests that there are differences in resource use and mortality by insurance status for patients with acute trauma even after adjustment for the severity of the injury. Given the acute nature of these conditions, such differences seem unlikely to be explained by patients' preferences. The observed differences in resource use and mortality occurred among patients with severe injuries as well as those cared for in the most well-equipped trauma centers. Taken together, these data raise the possibility that the excess mortality experienced by the uninsured may be avoided if their use of resources is increased to match that of insured patients. While health insurance may be a proxy for other socioeconomic factors, patients with Medicaid had outcomes that were more comparable to those of patients with private insurance, suggesting that the presence of health insurance conveys improved access to trauma-related care.

We examined three markers of trauma-related care that were available in hospital discharge abstract data. We chose the use of care in an intensive care unit and the use of an operative procedure as measures of trauma-related care because these interventions commonly are urgent or emergent. The finding that 81% of patients who received an operative procedure underwent surgery within 24 hours of admission supports that surgery was usually an urgent intervention. Although physical therapy is not emergent, it is central to the rehabilitation of patients with injuries.[19] The consistency of the observed differences in care across various subgroups suggests that the uninsured truly receive fewer resources.

Our results extend prior analyses of variation in care by payer in several ways. Although other studies have demonstrated differences in the use of resources by payer,[5-8] these studies usually have not been able to address the appropriateness of the observed variation. Hadley et al.[6] found higher rates of in-hospital mortality and lower rates of expensive inpatient procedures among uninsured vs privately insured adults, although these differences largely involved elective procedures

Braveman et al.[20] found that, among newborns with medical problems, those without insurance had shorter hospital stays than those with private insurance. Finally, Ayanian et al.[21] found that, among women with breast cancer, the uninsured presented with more advanced disease and had shorter survival, perhaps related to diminished access to medical care. We chose to examine trauma-related care because the interventions are emergent or urgent and, therefore, the link between care and mortality is more compelling.

There are several potential explanations of our findings. Because of the high costs associated with trauma care, some hospitals may treat the uninsured less aggressively. Several hospitals have stopped providing trauma care because of operating deficits.[22] The observed differences may reflect an effort to conserve costs among patients who potentially cannot pay.

Although we controlled for severity of injury using a validated algorithm, our data set does not include any physiologic data. It is possible that unmeasured differences in severity of injury explain the observed differences in resource use and mortality by insurance status. The Injury Severity Score may underestimate the severity of injury for patients with penetrating injuries. In the subgroup of patients with information about mechanism of injury, those who were uninsured were more likely to have a penetrating injury, and a penetrating injury was associated with a greater likelihood of death after Injury Severity Score was controlled. Although the ICD-9-CM system allows for the classification of mechanism of injury with its E codes, these codes were not recorded for most patients. For the subgroup of patients with information about the mechanism of injury, controlling for this information did not change the differences in resource use or mortality by payer. Similarly, our data set contains limited clinical data. It is possible that unmeasured differences in clinical circumstance explain the observed differences in resource use and mortality.

There are also several less plausible explanations of our findings. First, differences in the likelihood of hospitalization among patients with trauma might account for the observed treatment differences among patients who reached the hospital. However, if the uninsured were more likely to be admitted to the hospital and therefore less in need of trauma-related care, their mortality rates should be lower. Second, the uninsured were more likely to receive care in inner-city hospitals; these hospitals may be overcrowded,[23,24] and their resources may be more limited. However, the observed differences persisted after we controlled for hospital location.

An observational study cannot conclude that there is a causal relationship between the observed differences in trauma-related care and mortality. This study examined only hospital-based care. Differences in prehospital care would also affect mortality. Our data do not include individuals who died before reaching the hospital or who died in the emergency department.

In conclusion, among patients with acute trauma, those who were uninsured were less likely to receive several measures of trauma-related care and were more likely to suffer in-hospital mortality than were those with private insurance. If the use of fewer resources is contributing to the higher mortality rate for the uninsured, then an increase in resource use would narrow the mortality differences. Since trauma is the leading cause of death among young adults, future investigations should consider how to reduce the disparities in the allocation of trauma-related care and how to measure the effect of such efforts on mortality and other health-related outcomes. □

Acknowledgments

Dr Haas was the recipient of a National Research Service Award from the Agency for Health Care Policy and Research (5-F32-HS00025-03). The data for this project were provided through a grant from the Massachusetts Health Data Consortium and the Walnut Medical Charitable Trust.

We are grateful to Elliot Stone and staff of the Massachusetts Health Data Consortium for providing the data. We thank Robert Demling, MD, John Graham, PhD, and Ilana Lescohier, PhD, for helpful comments on an earlier version of the manuscript.

References

1. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1991. DHHS publication 91-50213.
2. Rice DP, MacKenzie EJ, Jones AS, et al. *Cost of Injury in the United States: A Report to Congress, 1989.* San Francisco, Calif: Institute for Health and Aging, University of California Injury Prevention Center; 1989.
3. Cayten CG, Stahl WM, Agarwal N, Murphy JG. Analysis of preventable deaths by mechanism of injury among 13,500 trauma admissions. *Ann Surg.* 1991;214:510–520.
4. Cales RH, Trunkey DD. Preventable trauma deaths: a review of trauma care systems development. *JAMA.* 1985;254:1059–1063.
5. Wenneker MB, Weissman JS, Epstein AM. The association of payor with utilization of cardiac procedures in Massachusetts. *JAMA.* 1990;264:1255–1260.
6. Hadley J, Steinberg EP, Feder J. Comparison of uninsured and privately insured hospital patients: condition on admission, resource use and outcome. *JAMA.* 1991;265: 374–379.
7. Weissman JS, Epstein AM. Case-mix and resource utilization by uninsured hospital patients in the Boston metropolitan area. *JAMA.* 1989;261:3572–3576.
8. Greenberg E, Chute C, Stukel T, et al. Social and economic factors in the choice of lung cancer treatment. *N Engl J Med.* 1988;318:612–617.
9. Johnson PA, Lee TH, Cook EF, Rouan GW, Goldman L. Effect of race on the presentation and management of patients with acute chest pain. *Ann Intern Med.* 1993;118:593–601.
10. *International Classification of Diseases, Ninth Revision, Clinical Modification.* Washington, DC: US Dept of Health and Human Services, 1991.
11. Committee on Medical Aspects of Automotive Safety. Rating the severity of tissue damage, I: the abbreviated scale. *JAMA.* 1971;215:277–280.
12. Committee on Injury Scaling. *The Abbreviated Injury Scale, 1985 Revision.* Morton Grove, Ill: American Association for Auto Medicine, 1985.
13. MacKenzie EJ, Steinwachs DM, Shankar B. Classifying trauma severity based on hospital discharge diagnoses: validation of an ICD-9-CM to AIS-85 conversion table. *Med Care.* 1989;27:412–422.
14. Baker S, O'Neill B, Haddon W, Long WB. The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma.* 1974;14:187–196.
15. Copes WS, Champion HR, Sacco WJ, Lawnick MM, Keast SL, Bain LW. The Injury Severity Score revisited. *J Trauma.* 1988;28:69–77.
16. Baker S, O'Neill B. The Injury Severity Score: an update. *J Trauma.* 1976;16:882–885.
17. Morris JA, MacKenzie EJ, Edelstein SL. The effect of pre-existing conditions on mortality in trauma patients. *JAMA.* 1990; 263:1942–1946.
18. *SAS/STAT User's Guide, Volume 2, Version 6.* 4th ed. Cary, NC: SAS Institute Inc; 1990.
19. Gordon DL, Reinstein L. Rehabilitation of the trauma patient. *Am Surg.* 1979;45:223–227.
20. Braveman PA, Egerter S, Bennett T, Showstack J. Differences in hospital resource allocation among sick newborns according to insurance coverage. *JAMA.* 1991;266:3300–3308.
21. Ayanian JZ, Kohler BA, Abe T, Epstein AM. The relation between health insurance coverage and clinical outcomes among women with breast cancer. *N Engl J Med.* 1993;329:326–331.
22. Hiatt JR, Kobayashi MR, Dees G, Ransom KJ. Trauma center closure; effects on an adjacent trauma center. *Am Surg.* 1991;57: 359–360.
23. Friedman E, Hagland MM, Hudson T, McNamara P. The sagging safety net; emergency departments on the brink of crisis. *Hospitals.* 1992;66:26–40.
24. Rice MF. Medical indigency and inner city hospital care: patient dumping, emergency care and public policy. *J Health Soc Policy.* 1989;1:1–29.

Comparing Hospitals That Perform Coronary Artery Bypass Surgery: The Effect of Outcome Measures and Data Sources

Arthur J. Hartz, MD, PhD, and Evelyn M. Kuhn, PhD

ABSTRACT

Objectives. The relative quality of hospital care often is judged by comparing risk-adjusted rates of adverse outcomes. This study evaluated whether hospital quality comparisons are affected by the choice of outcome and the use of administrative data instead of clinical data.

Methods. The data were collected from 2687 coronary artery bypass surgery patients from 17 hospitals. All patients were on Medicare. For 10 hospitals with 94 to 713 patients, risk-adjusted outcomes for death, major complications, and any complications were derived from a clinically rich database and an administrative database.

Results. The correlations between adjusted hospital rankings derived from the clinical and administrative databases were not significant: .48 for mortality, .21 for major complications, and −.14 for any complication. When only the clinical database was used, the correlation between risk-adjusted hospital rankings for mortality and major complications was .77 ($P < .01$) and the correlation between major complications and any complication was −.45.

Conclusions. These results suggest assessing quality of care by the use of administrative data may not be adequate and that quality assessment by the use of clinical data may depend greatly on the outcome chosen. (*Am J Public Health.* 1994;84: 1609–1614)

Introduction

Quality of care at hospitals that perform coronary artery bypass surgery frequently is assessed by comparing patient outcomes.[1–8] Although quality of care may be best assessed by evaluating the process of care,[9,10] assessing quality of care by evaluating outcomes also is important.[11] The reasons for using outcomes to evaluate quality of care are the following: (1) the goal of medical care is outcome; (2) the optimal process of care is not known; (3) the actual process of care usually is not completely recorded; (4) information needed to determine the process of care is more difficult to abstract and, therefore, more expensive to collect than information needed to determine risk-adjusted outcome; and (5) outcome is characterized and analyzed more easily than process.

Despite the widespread interest in comparing outcomes to measure quality of care, there is no concensus on how outcomes should be measured. Measurement of quality of care depends on the choice of outcome, the source of data used to determine outcome, and the data that should be used to take into account patient risk when measuring outcome. An ideal measurement that depends on risk-adjusted quality of life[5–8] and that uses comprehensive risk factors collected concurrently is too expensive to obtain for most assessments of quality of care.

The most easily obtained and, therefore, most widely used source to determine risk-adjusted outcomes is the Uniform Bill 82 (UB82) discharge abstract. This abstract includes principal and secondary patient diagnoses, some complications, discharge status, and patient demographic information. The most reliable and important outcome coded on the discharge abstract is patient mortality. This outcome is commonly used to assess the quality of care provided by hospitals that perform coronary artery bypass surgery.[1–3] Using mortality as a measure of outcome, however, will not reveal many errors in care and requires large numbers of patients in order to obtain adequate comparison of hospitals. To obtain a more sensitive measure of care, some studies have used patient complications coded on the discharge abstract.[12–21] Unfortunately, the data available from the discharge abstract are limited; there are frequent coding errors in the principal diagnosis and comorbidities,[14,19,23] and there may be even greater errors made when complications are coded.[14] Therefore, both the outcomes and the risk-adjustment procedure obtained from the discharge abstract may be flawed.

Even when better data from special data abstractions are available,[20–25] there is still the issue of which outcomes should be evaluated. For example, comparisons of coronary artery bypass surgery patients have used such outcomes as mortality rate,[1–3] length of stay,[26] readmissions,[27] and complications of surgery.[4,28–31] The present study evaluated whether choice of outcome and data sources affect the assessment of quality of care given by hospitals that perform coronary artery bypass surgery.

Arthur J. Hartz and Evelyn M. Kuhn are with the Department of Family and Community Medicine, Medical College of Wisconsin, Milwaukee.

Requests for reprints should be sent to Arthur J. Hartz, MD, PhD, Department of Family and Community Medicine, Medical College of Wisconsin, 8701 Watertown Plank Rd, Milwaukee, WI 53226.

This paper was accepted June 14, 1994.

TABLE 1—Patient Characteristics and Complication Rates from the Coronary Artery Bypass Surgery Abstract

Age	71.2 ± 6.3[a]
Female, %	33.8
Postprocedure length of stay for all patients	12.8 ± 12.3[a]
Major complications, %	
In-hospital mortality	8.0
Renal failure	4.2
Perioperative myocardial infarction	6.5
Long stay in the hospital and intensive care unit	6.6
Stroke	4.7
Readmission to intensive care unit	5.1
Other major complications[b]	2.5
Minor complications, %	
Other bleeding	20.2
Infection	14.3
On respirator for 48 hours or more	13.4
Return to surgery because of bleeding	6.9
Any complication, %	42.4

[a]Standard deviation.
[b]Bowel bleed, endocarditis, or pancreatitis.

TABLE 2—Relationship between Complication Rates[a] from Chart Review and from UB82 Administrative Data for Coronary Artery Bypass Surgery Patients

	Administrative/Clinical[b]		Clinical/Administrative[c]	
Complication	Proportion Identified	%	Proportion Identified	%
Renal failure	18/113	15.9	18/66	27.3
Endocarditis	0/8	0	. . .	. . .
Pancreatitis	0/21	0	0/2	0
Stroke	59/125	47.2	59/117	54.4
Bowel bleed	5/44	11.4	. . .	. . .
Any of the above	106/489	21.7	106/197	53.8
Major complications[d]	377/637	59.2	377/468	80.6
Other bleeding	146/568	25.7	146/195	74.9
Infection	46/385	12.0	46/55	83.6
Any complication[e]	589/1139	51.7	589/707	83.3

[a]All patients who had a given complication are included in calculating the complication rate regardless of whether they survived.
[b]The proportion of complications indentified during clinical review that also were coded in the UB82 administrative data.
[c]The proportion of complications identified in the UB82 administrative data that also were coded during clinical review.
[d]Major complications includes any complication listed above plus mortality and extended hospital stay.
[e]Any complication includes any complication listed above plus mortality and extended hospital stay.

Methods

For this study, we derived six risk-adjusted outcome-based measures of quality of care from two data sources and assessed the relationship among the measures.

Data Sources

Data were obtained on Medicare patients from all Wisconsin hospitals that perform coronary artery bypass surgery and from one hospital outside Wisconsin. Data from charts of all bypass surgery patients from the Wisconsin hospitals were abstracted by the Wisconsin Peer Review Organization. These data were obtained from charts the organization reviewed from April 1990 to June 1991. The development of the data collection form and data collection procedure for the Wisconsin patients has been described previously.[4] Data from the hospital outside Wisconsin were abstracted by hospital medical records personnel for patients hospitalized in fiscal year 1990. The same data abstraction form and instructions were used for these patients and the Wisconsin patients. Data for the out-of-state hospital and Wisconsin hospitals were comparable in that the hospitals had the same relationship between clinical risk factors and clinical outcomes and the same relationship between outcomes from the clinical data and outcomes from the administrative data.

Data were collected on 2687 patients from 17 hospitals, 15 of which were not-for-profit and 5 of which were teaching hospitals. Thirteen of the hospitals had more than 300 beds. Because the purpose of this study was to compare hospital rankings on the basis of different measures of patient outcome, however, the assessed hospitals had to have sample sizes large enough to obtain relatively stable estimates of complication rates. For this reason, only those nine hospitals that had at least 100 patients (one hospital had only 94 patients) were considered individually. A composite hospital with more than 100 patients was formed from the remaining hospitals. Patients from one smaller hospital were excluded from the analysis. This hospital had a rate of nonfatal major complications of only 2.1% compared with 15.7% for the average hospital and 10.5% for the hospital with the second lowest rate of nonfatal complications. In addition, personnel at this hospital acknowledged significant problems with the content of their medical records. For these reasons, we considered the error rate for the data from this hospital unacceptable for our study. Excluding patients from this hospital reduced the number of patients to 2641.

Patient Outcomes

The data elements were obtained from computerized UB82 billing data and from a coronary artery bypass surgery–specific abstract for which data were collected manually. The UB82 billing data included the patient's age and sex, up to five diagnoses of comorbidities or complications, three procedures, admission and discharge dates, number of days in the intensive care unit, and whether the patient was discharged alive. The coronary artery bypass surgery–specific data included information on any of the following major adverse outcomes: death, periprocedure myocardial infarction, renal impairment, stroke, readmission to the intensive care unit, endocarditis, bowel bleeds or necrosis, pancreatitis, or a long hospital and intensive care unit stay. Definitions for these outcomes have been presented previously.[4] A patient who died or had any of these adverse outcomes was considered to have a sustained "major adverse outcome" based on the coronary artery bypass surgery abstract data. "Any adverse outcome" was defined as a major adverse outcome or any of the following: the patient was on a respirator for 48 hours or more, developed an infection, returned to surgery because of bleeding, or required six or more units of blood, platelets, or plasma.

Adverse outcomes derived from the UB82 data were combinations of the following: death, postoperative length of stay greater than the 95th percentile (28 days), a major complication coded in the UB82 data, and any complication coded in the UB82 data. The major complications coded in the UB82 data were central nervous system complications (9970), urinary or renal complications (9975), endocarditis (4219), pancreatitis (5570), and postoperative shock (9980). Additional codes used for any complication were infection or inflammatory reaction due to internal prosthetic device, implant, or graft (9966); postoperative infection (9985 or 9993); and hemorrhage or hematoma complicating a procedure (9981).

In combinations of adverse outcomes from either the coronary artery bypass surgery abstract or the UB82 abstract, we always included major adverse outcomes when less major adverse outcomes also were included (e.g., every combination that included renal impairment also included death, and every combination that included return to surgery also included death and renal impairment).

TABLE 3—Hospital Complication Rates[a] for Coronary Artery Bypass Surgery Patients

	From Coronary Artery Bypass Surgery Abstract			From Administrative Data	
Hospital[b]	Major Complications	Any Complication	Hospital Mortality	Major Complications	Any Complication
1	14.7** (1)	62.9† (10)	4.3 (1)	12.9 (2)	25.0 (5)
2[c]	20.0 (2)	39.9** (4)	6.5 (3)	12.1*** (1)	18.4† (1)
3	21.3 (3)	35.1 (2)	6.4 (2)	30.9† (10)	33.0 (9)
4[c]	22.1 (4)	44.7 (5)	8.4 (7)	16.8 (6)	22.1 (2)
5	22.5 (5)	56.7*** (9)	7.5 (4)	13.3 (3)	23.3* (4)
6[c]	23.6 (6)	34.6† (1)	8.2 (5)	17.5 (7)	23.1* (3)
7	24.4 (7)	48.1 (7)	9.9 (10)	16.8 (5)	29.0 (7)
8[c]	26.1* (8)	39.7* (3)	8.3 (6)	18.1 (8)	27.6 (6)
9	26.2 (9)	46.6 (6)	8.7 (8)	15.5 (4)	31.1 (8)
10[c]	31.4*** (10)	53.8† (8)	9.9 (9)	26.9† (9)	43.9† (10)
All	23.7	42.4	8.0	17.4	26.3

[a]Ranking for a given complication is given in parentheses next to the complication rate.
[b]Hospitals are ordered according to rate of major complications.
[c]Hospitals with at least 190 patients.
*$P < .10$ for comparison of hospital rate with rate for all other patients; **$P < .05$; ***$P < .01$; †$P < .001$.

TABLE 4—Hospital Risk-Adjusted Relative Odds of a Complication[a] for Coronary Artery Bypass Surgery Patients

	From Bypass Surgery Abstract			From Administrative Data		
Hospital[b]	Hospital Mortality	Major Complications	Any Complication	Hospital Mortality	Major Complications	Any Complication
1	0.57 ((2)	0.54** (1)	2.61† (10)	0.49 (1)	0.68 (2)	0.95 (5)
2[c]	0.93 (4)	0.88 (2)	0.97 (4)	0.67 (2)	0.58*** (1)	0.58† (1)
3	0.87 (3)	0.94 (3)	0.67* (2)	0.69 (3)	1.94*** (10)	1.27 (9)
4[c]	1.07 (7)	1.01 (7)	1.22 (7)	1.11 (7)	0.93 (5)	0.76 (2)
5	1.25 (10)	1.12 (8)	2.19† (9)	0.90 (4)	0.72 (3)	0.84 (4)
6[c]	1.07 (6)	1.00 (6)	0.62† (1)	1.03 (6)	1.00 (6)	0.82* (3)
7	1.19 (8)	1.18 (9)	1.34 (8)	1.28 (10)	1.01 (7)	1.18 (6)
8[c]	1.21 (9)	1.23* (10)	0.91 (3)	1.16 (9)	1.20 (8)	1.25** (3)
9	0.98 (5)	0.96 (4)	1.08 (5)	0.96 (5)	0.76 (4)	1.22 (7)
10[c]	0.56** (1)	0.99 (5)	1.18 (6)	1.16 (8)	1.83† (9)	2.17† (10)

[a]Hospital ranking is given in parentheses next to the relative odds.
[b]Hospitals are ordered according to rate of major complications.
[c]Hospitals with at least 190 patients.
*$P < .10$; **$P < .05$; ***$P < .01$; †$P < .001$.

Methods to Adjust for Risk of Illness

Patient risk of illness was derived from information on the patient risk factors. The coronary artery bypass surgery abstract contained information on 35 patient risk factors that have been described previously.[4] Although ideally it is better to risk-adjust u ing only patient characteristics that are objectively identified and do not depend on physician judgment, physician management decisions sometimes provide patient information that is not present in the medical record or that is difficult for the nonspecialist to extract. We used physician decisions as proxy information for patient characteristics for two reasons: (1) to determine severely compromised cardiac function and (2) to characterize the anatomy of the coronary disease.

The patient's age and information on up to five diagnoses listed in the UB82 abstract were used to create the Charlson index, a risk of illness measure that has been used in many previous studies and has been found to be related to mortality rate for bypass surgery patients.[32-34] This index was tested for an association with patient outcome and used to risk-adjust hospital outcome. The Charlson index derived in this study may not have characterized a patient's risk of illness as well as a Charlson index derived when more diagnoses are available.

The other method we used to determine a patient's risk of illness for this study was logistic regression analysis. Stepwise logistic regression was used as described previously[4] to identify risk factors that have an independent association with a given outcome. We derived separate equations from risk factors limited to data elements from the UB82 abstract and from risk factors on the coronary artery bypass surgery abstract.

The risk of illness measure was used to derive a method to compare hospitals on the basis of patient outcome after taking into account patient risk. To calculate the relative odds of an adverse patient outcome at one hospital compared with all other hospitals, we used the logistic regression equation with an indicator variable for a given hospital. The relative odds of a complication for patients at a given hospital compared with patients in all other hospitals was e^b, where b is the regression coefficient of the indicator variable.[35]

TABLE 5—Correlations among Risk-Adjusted Outcome Measures for Coronary Artery Bypass Surgery Patients

	From Bypass Surgery Abstract			From Administrative Data		
	Hospital Mortality	Major Complications	Any Complication	Hospital Mortality	Major Complications	Any Complication
Coronary artery bypass surgery						
Major complications	0.77***					
Any complication	−0.22	−0.45 (0.38)[a]				
Administrative data						
Hospital mortality	0.48	0.83***	−0.38			
Major complications	−0.31	0.21	−0.45	0.22		
Any complication	−0.46	0.17	−0.14	0.37	0.75**	
Unadjusted outcome[b]	0.33	0.62*	0.94†	0.95†	0.98†	0.99†

[a]Correlation after eliminating an extreme outlier.
[b]Correlations of the adjusted rates with the corresponding unadjusted rates.
*$P < .10$; **$P < .05$; ***$P < .01$; †$P < .001$.

Results

Patient characteristics and complication rates are shown in Table 1. Because the patients in the study were on Medicare, almost all were older than 64 years, and the mean age was 71. The average length of stay after the procedure was approximately 13 days. The most common major complications were mortality (8.0%), long length of stay in the hospital or intensive care unit (6.6%) and postoperative myocardial infarction (6.5%).

The relationship between the complications identified by the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), codes and complications identified from the UB82 abstract are shown in Table 2. With the exception of stroke, fewer than 26% of the complications identified on the coronary artery bypass surgery abstract were coded on the UB82 abstract. The UB82 abstract was moderately sensitive (50% to 60%) at identifying the existence of major complications or any complication primarily because the sensitivity of the form for detecting mortality or long stay was considered to be 100%. To detect at least one of the other major complications, the sensitivity was only 21.7%.

One reason for the low sensitivity of the UB82 abstract in identifying complications other than mortality or long stay was that there was room for only five diagnoses and complications on the form used by most hospitals. To determine whether space limitations contributed to the low sensitivity, we examined the data from UB82 abstracts from one hospital that listed nine diagnoses on its form. For this hospital, the sensitivity of the UB82 abstract for detecting major complications other than mortality increased from 21.7% with five diagnoses to 38.4% with nine diagnoses.

The right column of Table 2—"Clinical/Administrative"—shows the percentage of complications coded on the UB82 abstracts that were identified on the coronary artery bypass surgery abstracts. These percentages were much greater than the percentages under "Administrative/Clinical"; for major complications, infection, and any complications, the "Clinical/Administrative" percentage were greater than 80%. The percentage was not higher because of errors or limitations of the coronary artery bypass surgery abstract and, more important, because the ICD-9-CM code is not specific (i.e., the same code refers to many complications, both major and minor).

The rates for each adverse outcome of the 10 hospitals are shown in Table 3. The hospitals are ordered by the rate of major complications, which varied significantly among the hospitals ($P = .02$ by the chi-square test for contingency tables). There was nearly a twofold variation in the major complication rate, and two of the hospitals had rates that differed significantly at the $P < .05$ level from the rates for all other patients combined. There also was substantial variation in the rates for any complication ($P < .0001$); five hospitals had rates of any complication that differed from the rates of other hospitals at the $P < .05$ level of significance. Some rankings were similar for different outcomes, and some rankings differed greatly. One hospital had the lowest rate of major complications (a ranking of 1) and the highest rate of any complication (a ranking of 10) even though 56% of the patients who had any complication had a major complication.

Hospital rankings adjusted for patient risk of illness are shown in Table 4. The hospitals are listed in the same order as in Table 3 to make it possible to compare unadjusted and adjusted rankings. Some hospitals were substantially affected by the adjustment procedure for some outcomes. The most extreme example is that the hospital with the ninth highest unadjusted morality rate had the lowest rate after adjusting for the factors in the coronary artery bypass surgery abstract. Another example is that the hospital with the highest unadjusted rate of major complications had the fifth highest adjusted risk of major complications.

The correlations among the risk-adjusted outcome-based measures of quality of care are shown in Table 5. The risk-adjusted outcomes from the UB82 abstracts had nonsignificant correlations with the risk-adjusted outcomes from the coronary artery bypass surgery abstracts. Two of the risk-adjusted outcomes derived from the coronary artery bypass surgery abstracts were significantly correlated—death and major complications ($r = .77$, $P < .01$)—but the risk of all complications was not significantly correlated with these two risks.

The last row of Table 5 shows the correlation between corresponding unadjusted and adjusted outcome measures. Mortality and the rate of major complications were substantially affected by adjustment for variables in the coronary artery bypass surgery abstracts. Adjusting for the information from the UB82 abstracts made little difference in hospital assessment (i.e., all of the correlations were greater than .90). Correlations between unadjusted rates and rates adjusted with the Charlson index were greater than .97.

Table 5 does not show the correlation between the risk-adjusted mortality rate and the rate of major nonfatal complications. The correlation for these rates was .31 for all hospitals and .79, after eliminating one hospital. Using a bootstrap procedure to test the significance of the maximum correlation obtained by choosing 9 of 10 points, we found that the P value for a correlation of .79 was .035.

Discussion

The findings of this study suggest that the results of outcome-based assessments of quality of care at hospitals that perform coronary artery bypass surgery varies depending on the source of data and the outcome measure used. Neither the outcome measures nor the risk factors from the UB82 abstracts were precise or accurate enough to derive quality measures that reflected those derived from the coronary artery bypass surgery abstracts. Risk adjustment for mortality was not adequate with the UB82 abstract data, and complications frequently were not recorded in these data. Even when as many as nine ICD-9-CM codes were captured, fewer than 40% of the complications from the coronary artery bypass surgery abstracts were captured with the UB82 abstracts. The codes that were recorded frequently were not specific enough to distinguish major from minor complications.

As found in previous mortality studies,[4,12] risk adjustment with UB82 in our study data was not nearly as effective as risk adjustment with coronary artery bypass surgery abstract data. Risk adjustment did not affect rankings for most hospitals because most hospitals had patients at a similar level of risk; however, risk adjustment did make a substantial difference for some hospitals.

Choice of outcome also affected measurement of quality of care. Although there seemed to be a relationship at most hospitals between the risk-adjusted rates of mortality and major complications, there was no statistical evidence that hospital rates of any complication were associated with hospital rates of major complications or mortality.

The major advantage of this study was that it employed a common, comprehensive data set from several hospitals. The major disadvantage was that neither the sample sizes per hospital nor the number of hospitals was large. With larger sample sizes, we would have been able to estimate more precisely the strength of associations between the different measures of quality of care. A second disadvantage is that we did not have any standard measure of quality of care, such as postdischarge quality-adjusted survival. Without such a standard, it was not possible to determine which outcome measures provided the best information about the quality of care. A third disadvantage is that we were not able to evaluate whether the quality of data or the types of missing data differed by hospital type.

Recent outcome-based measures of quality of care have been sophisticated and varied.[1-21,25-31] Hospital comparisons, however, generally are restricted to one outcome. Our results suggest that performance across outcomes is not uniform. It is possible that optimum measurement of quality of care may require the examination of several outcomes. Each outcome-based measure of quality of care may reflect different aspects of the process of care, and hospitals should be informed about their performance regarding all aspects of care that can be evaluated adequately. □

Acknowledgment

This work was supported in part by grants HS06285-02 and HS-06353 from the Agency for Health Care Policy Research.

References

1. Hannan EL, Kilburn H Jr, O'Donnell JF, Lukacik G, Shields EP. Adult open heart surgery in New York state: an analysis of risk factors and mortality rates. *JAMA*. 1990;264:2768–2774.
2. O'Connor GT, Plume SK, Olmstead EM, et al. A regional prospective study of in-hospital mortality associated with coronary artery bypass grafting. *JAMA*. 1991;266:803–809.
3. Williams SV, Nash DB, Goldfarb N. Differences in mortality from coronary artery bypass graft surgery at five teaching hospitals. *JAMA*. 1991;266:810–815.
4. Hartz AJ, Kuhn EM, Kayser KL, Pryor DP, Green R, Rimm AA. Assessing providers of coronary revascularization: a method for peer review organizations. *Am J Public Health*. 1992;82:1631–1640.
5. Wortman PM, Yeaton WH. Cumulating quality of life results in controlled trials of coronary artery bypass graft surgery. *Controlled Clin Trials*. 1985;6:289–305.
6. Bunzel B, Eckersberger F. Change in the quality of life following aortocoronary bypass and heart valve replacement: assessment of subjective surgical success. *Thorac Cardiovasc Surg*. 1987;35:242–247.
7. Mayou R, Bryant B. Quality of life after coronary artery surgery. *Q J Med*. 1987;62:239–248.
8. King KB, Porter LA, Norsen LH, Reis HT. Patient perceptions of quality of life after coronary artery surgery: was it worth it? *Res Nurs Health*. 1992;15:327–334.
9. Berwick DM, Knapp MG. Quality assurance in hospitals strategies for assessment and implementation. In: Graham NO, ed. *Theory and Practice for Measuring Health Care Quality*. 2nd ed. Rockville, Md: Aspen Publishers; 1990:118–131.
10. McAuliffe WE. Measuring the quality of medical care: process vs. outcome. *Milbank Mem Fund Q/Health & Soc*. 1979;57(1):118–152.
11. Donabedian A. Defining and measuring the quality of health care. In: Wenzel RP, ed. *Assessing Quality Health Care Perspectives for Clinicians*. Baltimore, Md: Williams & Wilkins; 1992:41–64.
12. Hannan EL, Kilburn H, Lindsey ML, Lewis R. Clinical versus administrative data bases for CABG surgery: does it matter? *Med Care*. 1992;30:892–907.
13. Hannan EL, Bernard HR, O'Donnell JF, Kilburn H. A methodology for targeting hospital cases for quality of care record reviews. *Am J Public Health*. 1989;79:430–436.
14. Iezzoni LI. Using administrative diagnostic data to assess the quality of hospital care: pitfalls and potential of ICD-9-CM. *Int J Technol Assess Health Care*. 1990;6:272–281.
15. Quam L, Ellis LBM, Venus P, Clouse J, Taylor CG, Leatherman S. Using claims data for epidemiologic research: the concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. *Med Care*. 1993;31:498–507.
16. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias: does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA*. 1992;267:2197–2203.
17. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. *Med Care*. 1994;32:81–90.
18. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. *Ann Intern Med*. 1993;119:844–850.
19. Wennberg JE, Roos N, Sola L, Schori A, Jaffe R. Use of claims data systems to evaluate health care outcomes: mortality and reoperation following prostatectomy. *JAMA*. 1987;257:933–936.
20. DesHarnais S. Current uses of large data sets to assess the quality of providers. *Int J Technol Assess Health Care*. 1990;6:229–238.
21. Luft HS, Garnick DW, Mark DH, Peltzman DJ, Phibbs CS, Lichtenberg E, McPhee SJ. Does quality influence choice of hospital? *JAMA*. 1990;263:2899–2906.
22. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, Hsia DC. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health*. 1992;82:243–248.
23. Brewster AC, Karlin BG, Hyde LA, Jacobs CM, Bradbury RC, Young MC. MEDISGROUPS: a clinically based approach to classifying hospital patients at admission. *Inquiry*. 1985;22:377–387.
24. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification. *Crit Care Med*. 1985;13:818–829.
25. Iezzoni LI, Restuccia JD, Shwartz M, Schaumburg D, Coffman GA, Kreger BE, Butterly JR, Selker HP. The utility of severity of illness information in assessing the quality of hospital care: the role of the clinical trajectory. *Med Care*. 1992;30:428–444.
26. Weintraub WS, Jones EL, Craver J, Guy-

ton R, Cohen C. Determinants of prolonged length of hospital stay after coronary bypass surgery. *Circulation.* 1989;80: 276–284.

27. Rehospitalization by geographic area for aged Medicare beneficiaries: selected procedures, 1986–87. *Health Care Financing Administration Special Report.* Baltimore Md: Health Care Financing Administration, Office of Research and Demonstrations; June 1990. HCFA publication 03303.
28. Geraci JM, Rosen AK, Ash AS, McNiff KJ, Moskowitz MA. Predicting the occurrence of adverse events after coronary artery bypass surgery. *Ann Intern Med.* 1993;118:18–24.
29. Higgins TL, Estafanous FG, Loop FD, Beck GJ, Blum JM, Paranandi L. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients: a clinical severity score. *JAMA.* 1992;267:2344–2348.
30. Hammermeister KE, Burchfiel C, Johnson R, Grover FL. Identification of patients at greatest risk for developing major complications at cardiac surgery. *Circulation.* 1990; 82:IV380–IV389.
31. Huysmans HA, van Ark E. Predictors of perioperative mortality, morbidity, and late quality of life in coronary bypass surgery. *Eur Heart J.* 1989;10(suppl H):10–12.
32. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis.* 1987;40:373–383.
33. Roos LL, Sharp SM, Cohen MM, Wajda A. Risk adjustment in claims-based research: the search for efficient approaches. *J Clin Epidemiol.* 1989;42:1193–1206.
34. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol.* 1992;45:613–619.
35. Schlesselman JJ. Case-control studies: design, conduct, analysis. [In: *Monographs in Epidemiology and Biostatistics.*] New York, NY: Oxford University Press; 1982.

Call for Abstracts for International Symposium on Epidemiology in Occupational Health

A call for abstracts has been issued for the 11th International Symposium on Epidemiology in Occupational Health, which will be held September 5–8, 1995, in Noordwijkerhout, near Amsterdam, the Netherlands. The symposium, which will offer an international platform for the exchange of knowledge on occupational epidemiology, will include plenary and parallel sessions related to the following themes:

- Occupational epidemiology methods
- Occupational surveillance methods, evaluation, and ethical aspects
- Occupational diseases, injuries, and cancer
- Neurological, reproductive, and developmental effects
- Respiratory diseases
- Musculoskeletal disorders
- Work-related stress
- Exposure assessment in occupational epidemiology
- Biomarkers
- Communicable diseases at work

The symposium is being organized by the Department of Epidemiology and Public Health and Department of Air Quality at the Agricultural University, Wageningen, and the Department of Public Health at Erasmus University, Rotterdam, on behalf of the Scientific Committee on Occupational Epidemiology of the International Committee on Occupational Health.

The deadline for submission of abstracts is *January 31, 1995.* For more information, contact the conference secretariat: Susan Peelen, MSc, Department of Epidemiology and Public Health, PO Box 238, 6700 AE Wageningen, the Netherlands; tel +31 8370 84124; fax +31 8370 82782. E-mail: susan.peelen;@medew.hegl.wau.nl.

Rates, Patterns, Causes, and Costs of Hospitalization of Nursing Home Residents: A Population-Based Study

William H. Barker, MD, James G. Zimmer, MD, W. Jackson Hall, PhD, Brian C. Ruff, Charlene B. Freundlich, and Gerald M. Eggert, PhD

ABSTRACT

Objectives. Hospitalization of nursing home residents is a growing, poorly defined problem. The purposes of this study were to define rates, patterns, costs, and outcomes of hospitalizations from nursing homes and to consider implications for reducing this problem as part of health care reform.

Methods. Communitywide nursing home utilization review and hospital discharge data were used to define retrospectively a cohort of 2120 patients newly admitted to nursing homes; these patients were followed for 2 years to identify all hospitalizations. Resident characteristics were analyzed for predictors of hospitalization. Charges and outcomes were compared with hospitalization of community-dwelling elders.

Results. Hospitalization rates were strikingly higher for intermediate vs skilled levels of care (566 and 346 per 1000 resident years, respectively). Approximately 40% of all hospitalizations occurred within 3 months of admission. No strong predictors were identified. Length of stay, charges, and mortality rates were higher than for hospitalizations from the community.

Conclusions. Hospitalizations from nursing homes are not easily predicted but may in large part be prevented through health care reforms that integrate acute and long-term care. (*Am J Public Health.* 1994;84:1615–1620)

Introduction

The elderly population experiences not only disproportionately high rates of admission to hospitals and nursing homes but high rates of transfer between these settings. Transfer from hospitals to nursing homes has been extensively studied during the past 2 decades.[1,2] However, while there have been studies involving selected institutions,[3–8] there have been virtually no population-based studies of transfers from nursing homes to hospitals in the United States. The magnitude of the problem is substantial: the 1977 and 1985 National Nursing Home Surveys reported, respectively, 339 500 and 430 200 discharges to hospitals, resulting in estimated annual ratios of 242 and 265 hospitalizations per 1000 long-term care beds.[9,10] Such hospitalizations are costly[11] and potentially detrimental to frail elderly people[12,13]; hence, they are important to health services research and policymaking.

The dearth of population-based studies of hospitalization from nursing homes may be ascribed in part to the lack of data systems that link hospital and nursing home use. This, in turn, is a reflection of the historic separation between financing and organization of acute and long-term care in the United States.[14,15] The present study used unique computerized data sets available on all nursing home admissions and hospitalizations in the early 1980s for a defined geopolitical area. The purposes were to determine rates, predictors, causes, and costs of hospitalization of nursing home residents and to consider strategies for reducing these occurrences as part of national health care reform.

Materials and Methods

Setting

The setting for the study, Monroe County, New York, is an urban-suburban-rural jurisdiction centered on the city of Rochester. Table 1 compares selected characteristics for Monroe County and the country at large in the early 1980s. The ratio of general hospital beds per 1000 persons was somewhat smaller in Monroe County, while the ratio of nursing home beds per 1000 elderly persons (those more than 65 years of age) was slightly higher. At the time of this study, approximately 30% and 70% of nursing home beds were classified as intermediate care and skilled care, respectively, in both Monroe County and the country as a whole.[16] Specifically, there were 33 licensed nursing homes in Monroe County in 1982, all of which included skilled care beds (approximately 3600) and 9 of which included intermediate care beds (approximately 1400). (Intermediate and skilled care facilities are defined in the Appendix, and the differences between the two are described. With passage of the 1987 Omnibus Budget Reconciliation Act nursing home reform legislation, the skilled nursing facility and intermediate care

William H. Barker, James G. Zimmer, and Charlene B. Freundlich are with the Department of Community and Preventive Medicine, and W. Jackson Hall is with the Department of Biostatistics, University of Rochester Medical Center, Rochester, NY. Brian C. Ruff and Gerald M. Eggert are with the Monroe County, New York, Long Term Care Program.

Requests for reprints should be sent to William H. Barker, MD, Department of Community and Preventive Medicine, Box 644, University of Rochester Medical Center, Rochester, NY 14642.

This paper was accepted March 30, 1994.

TABLE 1—Population and Hospital and Nursing Home Bed Provision: Monroe County, New York, and the United States

	Monroe County	United States
Population[a]	702 000	226×10^6
Population >65 years of age[a] (%)	77 000 (11)	25.5×10^6 (11)
Hospital beds per 1000[b]	3.6	4.5
Nursing home beds[b]	4946	1.4×10^6
Nursing home beds per 1000 people >65 years of age[b]	65	60
Skilled nursing beds, %	70	66
Intermediate care beds, %	30	34

[a]1980 US census.
[b]1982 statistics, Finger Lakes Health Systems Agency and reference 16.

TABLE 2—Hospitalizations among Cohorts of Nursing Home Admissions during the 2-Year Follow-Up Period

	Skilled Nursing Facility Cohort	Intermediate Care Facility Cohort
Nursing home admissions, no.	1700	420
Any hospitalizations, no. (%)	451 (26.5)	175 (41.7)*
1 hospitalization	315 (18.5)	119 (28.3)*
2 hospitalizations	95 (5.6)	44 (10.5)*
3+ hospitalizations	41 (2.4)	12 (2.9)
Total hospitalizations	647	245
Total resident-years	1869	433
Hospitalizations per 1000 resident-years	346	566*

*$P < .001$ (for difference between rates for skilled nursing facility and intermediate care facility residents).

facility designations have been replaced by the single designation of residential health care facility for all nursing home beds.[17])

Design and Data

A retrospective cohort design was used in which all patients newly admitted to nursing homes in the county during a single calendar year (1982) were identified and followed for 2 years or until death to ascertain and analyze all episodes of transfer to acute hospitals.

The study population consisted of 2120 nursing home patients admitted between January 1 and December 31, 1982 (1700 and 420, respectively, to skilled and intermediate levels of care). The average length of stay during the 2-year period of follow-up was 402 days for the cohort of skilled nursing facility residents and 376 days for the cohort of intermediate care facility residents.

The principal source for identifying and characterizing the study cohort and for ascertaining hospitalizations was the computerized utilization review file maintained by the Monroe County Long Term Care Program, one of the first community-wide long-term care case management agencies in the country. The program was responsible for tracking all nursing home admissions and discharges in the county between 1977 and 1985.[18] Physical function and mental and behavioral status data were based on the Disability Measurement Scale (DMS-1) utilization review scale in routine use in nursing homes in New York State in the 1970s and early 1980s.[19] Baseline chronic medical conditions for cohort members were obtained by hand from utilization review forms maintained by the Genesee Valley Medical Foundation, a nonprofit agency that conducts, under contract, utilization reviews in nursing homes in Monroe County. These conditions were classified into broad disease categories developed for an earlier study of nursing home utilization in Monroe County.[20]

Discharge diagnoses, lengths of stay, and charges and outcomes of hospitalizations were obtained from the uniform billing forms routinely used by all hospitals in New York State. These forms, obtained through cooperative arrangements with all eight acute care hospitals in the county, accounted for all hospitalizations in the cohort. For comparison with hospitalizations among the community-dwelling population more than 65 years of age, computerized uniform billing form data were obtained from the Rochester Area Hospital Experimental Payment Program, which was in effect during the 1980s.[21] (The uniform billing form provides total dollar charges but not costs; therefore, we report comparative charges data. Because the hospitalizations from nursing homes and from the community involve the same hospitals, the relation of costs to charges was presumed to be similar for the two groups.)

Analysis

Overall rates of hospitalization and rates per 1000 resident-years were computed. Analyses for predictors of hospitalization were conducted separately for skilled nursing facility and intermediate care facility cohorts. Univariate analyses (with chi-square tests for significance) were performed to determine associations between demographic, medical, and functional status characteristics present at admission to nursing home and subsequent hospitalization. Stepwise multiple logistic regression analyses were performed to determine combinations of predictors of hospitalization (including any interactions) among admission characteristics; analyses were repeated for prediction of multiple hospitalizations. Distribution of hospitalizations by principal discharge diagnosis, length of stay, charges, and status at discharge (dead or alive) were computed for all hospitalizations and for those within each of the three age subgroups: 65 through 74, 75 through 84, and 85+ years. These data were compared with information on hospitalizations of people more than 65 years old living in the community.

Results

A total of 892 hospitalizations occurred among the 2120 subjects (a computed rate of 387 hospitalizations per 1000 resident-years). Table 2 summarizes overall frequency and rates of hospitalization for the skilled nursing facility and intermediate care facility cohorts. The percentages of residents experiencing at least one hospitalization during the 2-year follow-up period differed markedly by

level of care: 26.5% for skilled nursing facility residents vs 41.7% for intermediate care facility residents. Some 8% of skilled nursing facility and 13% of intermediate care facility residents experienced multiple discharges to a hospital. After adjustment for time at risk, the difference between the two cohorts persisted, with rates of 346 (skilled nursing facilities) and 566 (intermediate care facilities) total hospitalizations per 1000 resident-years. Cumulative percentages of hospitalization over time, for those ever hospitalized, are shown in Table 3. Graphs (not shown) were quite similar for the two groups, with the skilled nursing facility curve rising slightly faster (i.e., having a slightly heavier concentration of hospitalizations in the first 60 days). Approximately 40% of hospitalizations occurred within the first 90 days, and roughly three quarters within the first year, in each group.

TABLE 3—Number, Percentage, and Cumulative Percentage of Initial Hospitalizations among Admission Cohorts of Nursing Home Residents, by Time Intervals from Day of Nursing Home Admission

Time Interval from Admission, d	Skilled Nursing Facility Cohort			Intermediate Care Facility Cohort		
	No.	%	Cumulative %	No.	%	Cumulative %
1–30	81	18.0	18.0	25	14.3	14.3
31–60	51	11.3	29.3	19	10.9	25.2
61–90	41	9.1	38.4	20	11.4	36.6
91–180	84	18.6	57.0	34	19.4	56.0
181–360	87	20.0	77.0	25	14.3	70.3
361–540	57	12.6	89.6	25	14.3	84.6
541–730	50	11.1	99.7	27	15.4	100.0

TABLE 4—Percentage of Skilled Care and Intermediate Care Nursing Home Residents with One or More Hospitalizations during the 2-Year Follow-Up, by Selected Admission Characteristics

	No. with Characteristic (% with Hospitalization)	
	Skilled Nursing Facility Admissions	Intermediate Care Facility Admissions
All admissions	1700 (26.5)	420 (41.7)
Admitted from		
Home	304 (23.4)	171 (42.1)
Hospital	1160 (27.1)	140 (42.9)
Other institution	236 (26.3)	109 (39.5)
Resident age, y		
<65	92 (17.4)	30 (36.7)
65–74	254 (24.8)	50 (52.0)
75–84	603 (28.9)	150 (42.7)
85+	751 (26.2)	190 (38.9)
Resident sex		
Female	1264 (25.4)	323 (44.0)
Male	436 (29.9)*	97 (37.1)
Chronic medical condition		
Arthritis	367 (27.8)	131 (43.5)
Cardiac	695 (29.1)	197 (43.7)
Cerebrovascular	312 (28.2)	47 (40.4)
Chronic obstructive pulmonary disease	120 (27.5)	39 (53.8)
Dementia	464 (27.8)	79 (38.0)
Depression	84 (35.7)	46 (50.0)
Diabetes	232 (30.6)	54 (44.4)
Gastrointestinal	109 (33.9)	27 (48.1)
Hip fracture	237 (25.7)	34 (47.1)
On-site medical staff		
Yes	409 (21.8)	168 (44.6)
No	1290 (28.3)*	252 (39.7)

*$P < .05$.

Predictors

Tables 4 and 5 examine the association between individual baseline admission characteristics and subsequent hospitalization. For each category of baseline characteristics, hospitalization rates were compared within the skilled nursing facility and intermediate care facility cohorts. Location from which subjects were initially admitted to a nursing home, whether hospital, home, or another institution, was not associated with greater or lesser risk of subsequent hospitalization in either cohort. There was also little difference within the cohorts in risk of hospitalization among age groups, except for slightly lower percentages for the small number of residents less than 65 years of age. Men had higher rates than women in the skilled nursing facility cohort ($P < .05$). Residents with diabetes, gastrointestinal disease, or depression at the time of admission had somewhat higher risks of subsequent hospitalization in the skilled nursing facility cohort, while intermediate care facility residents with cancer, chronic obstructive pulmonary disease, seizure disorder, or depression had somewhat higher risks of hospitalization; however, chi-square tests comparing each of these rates with those among members of the respective cohorts with other baseline chronic conditions showed P values greater than .10. Within the skilled nursing facility cohort, residents in several large nursing homes with full-time medical staffs experienced lower subsequent rates of hospitalization ($P < .05$). Such relationships were not observed among the intermediate care facility subjects in the same institutions. Analyses by baseline functional status (Table 5), restricted to skilled nursing facility residents, revealed a lower rate of hospitalization for the small number who were totally bedbound ($P < .05$).

In summary, in the skilled nursing facilities, only male gender and absence of on-site medical staff were statistically significant predictors of hospitalization, and even then the differences in rates were not large. Being bedbound was associated with a significant lowering of risk. Source of admission, age group, and chronic medical conditions were not significant. For the intermediate care facilities, there were no significant predictors.

TABLE 5—Percentage of Skilled Care Nursing Home Residents with One or More Hospitalizations during the 2-Year Follow-Up, by Selected Physical, Mental, and Behavior Functions

Functional Status	No. with Characteristic	% with Hospitalization
Ambulation[a]		
1	585	26.6
2	1052	28.0
3 (bedbound)	64	11.0*
Urinary continence[b]		
1	717	23.2
2	288	28.2
3	581	29.6
4 (catheter)	164	23.2
Bathing[a]		
1	117	28.2
2	1279	26.3
3	304	22.7
Assaultive–abusive		
1	1565	26.8
2	91	27.5
3	44	15.9
Wandering		
1	1480	26.3
2	109	27.7
3	111	28.8

Note. Functional status, based on admission assessment data from the DMS-1 instrument,[19] was not analyzed for intermediate care facility residents because most were independent or minimally impaired in functions at the time of the admission assessment.

[a]1 = independent, unimpaired; 2 = partial dependence or impairment; 3 = total dependence or impairment.

[b]1 = no incontinence; 2 = intermittent incontinence; 3 = always incontinent; 4 = indwelling catheter.

*$P < .05$.

Stepwise multiple logistic regression analyses failed to reveal single characteristics or combinations of characteristics, including those listed in Tables 4 and 5, that were any more predictive of hospitalization than the limited associations found in univariate analyses. We were likewise unable to identify any combinations of characteristics predictive of (1) hospitalizations that occurred early vs later in the nursing home stay or (2) multiple hospitalizations.

Causes, Charges, and Outcomes

In Table 6, the most common principal discharge diagnosis groups are compared between hospitalizations that occurred among older persons in nursing homes (n = 808) and those living in the community (n = 24 891). Because the distributions by discharge diagnoses were very similar for hospitalizations from the intermediate care facility and skilled nursing facility cohorts, these data have been combined in the analysis. Noteworthy are the significantly greater proportions of cardiovascular and neoplastic disease and cataract surgery hospitalizations among community-dwelling elderly people and of pneumonia and hip fracture hospitalizations among nursing home residents. Overall average lengths of stay in the hospital were approximately 1 day longer (14.4 vs 13.2) and average charges slightly higher for elderly persons admitted from nursing homes than for those from the community (Table 7). These figures showed interesting reverse trends when analyzed for age subgroups, with the younger nursing home residents and the more elderly community-dwelling persons experiencing the longer and costlier hospitalizations within their respective populations. Within each age subgroup, a higher percentage of patients from nursing homes were dead at discharge from the hospital.

Discussion

A number of excellent literature reviews have focused on acute medical care problems of nursing home residents, with particular emphasis on transfers to acute hospitals.[22,23] The literature is composed largely of short-term studies conducted from the limited perspective of one or more selected hospitals, nursing homes, or patient subsets; thus, it is difficult to apply to an understanding of the problem from the perspective of a general population of nursing home residents. By contrast, the present study, using data available from two concurrent communitywide demonstrations,[18,21] reported rates and patterns of acute hospitalization over a 2-year period among a cohort of all patients newly admitted to nursing homes within a defined population. In brief, the study found a rate of hospitalization that was high relative to rates computed from National Nursing Home Surveys, concentrated within the first 3 months of admission to the nursing home, and most strongly associated with the level of care to which residents were assigned on admission. Within each of the two broad levels of care (skilled and intermediate), resident age, sex, medical conditions, and functional parameters at admission, as well as prior locus (home, hospital, other), were generally not strongly associated with risk of subsequent hospitalization, while the presence of on-site medical staff in large institutions was associated with a lower risk for skilled nursing facility residents. Finally, cost of care, as reflected in hospital charges, and mortality rates were somewhat higher, on average, for hospitalizations of nursing home residents than for hospitalizations of community-dwelling elderly people. How should these findings be interpreted, and what are the policy implications for health service delivery for nursing home residents?

The observed strikingly higher rate of hospitalization in the present study (387 per 1000 resident-years), in contrast with ratios derived from the 1977 and 1985 National Nursing Home Surveys data (242 and 265 per 1000 long-term beds per year), may in part reflect a regional tendency toward higher rates of use of acute hospitals by nursing home residents. Given the relatively low hospital bed to population ratio in Monroe County (Table 1) and the well-recognized overall lower rates of hospital admission in the county throughout the 1980s,[24] this is probably not a major factor. The more likely explanation lies in the different population frames of reference for the respective studies. The Monroe County study population, an admission cohort of nursing home residents, would include a higher percentage of relatively short-stay, potentially unstable subjects at higher risk of acute medical decline and transfer to hospital, while the National Nursing Home Surveys data, based in substantial part on a prevalent population of nursing home residents, would include a greater proportion of long-stay, more stable subjects.[25]

There are several possible reasons for the interesting, seemingly paradoxical finding of higher rates of hospitalization among people admitted to lower levels of nursing home care. A leading possibility is that the lesser degree of medical and nursing supervision for patients at lower levels of chronic care leads to a need to transfer unstable residents to hospitals more often than would be necessary from better staffed skilled nursing facilities (see Appendix). In addition, given the greater prevalence of compromising physical and mental disability among skilled care residents, it may be that heroic medical intervention, including transfer to a hospital in instances of acute illness, is withheld more often among these residents. The withholding of life-saving interventions, including hospitalizations, per the ex-

pressed wishes of older persons in nursing homes has been increasingly discussed and formalized in recent years.[26–28]

The failure to more fully identify individual characteristics or patterns of characteristics predictive of which nursing home residents are most at risk of hospitalization within either the skilled nursing facility or intermediate care facility level of care may be attributable to limitations in the utilization review data used in this study. While utilization review at admission is intended primarily for classifying residents by level of need for assistance with activities of daily living within the nursing home, it is readily acknowledged that baseline utilization review data on medical conditions may be incomplete and do not permit classification of conditions by medical severity,[29] which is clearly a most important factor in risk of hospitalization.[30]

Nursing home residents, like community-dwelling elderly persons,[31] had relatively high overall risks of hospitalization as well as rehospitalization. The somewhat higher average charges and mortality rates and higher proportion of respiratory infections and fractures involved in hospitalizations of those from the nursing home cohort in comparison with those from the community are similar to findings reported elsewhere.[3–6] Recognizing that many nursing home hospitalizations might be avoided by applying relatively unsophisticated acute services on site (e.g., intravenous antibiotics), various earlier studies have estimated that as many as 40% to 50% of these hospitalizations might be preventable.[8,32] Given the substantial costs and frequency of hospitalization of nursing home residents and evidence that these factors may be increasing since implementation of the Medicare Prospective Payment System,[2,22] reducing such hospitalizations becomes a serious policy consideration. (Hospital discharges from Monroe County nursing homes ranged between 900 and 1200 from 1986 to 1990 [L. Varricchio, Finger Lakes Health Systems Agency, written communication, August 1993].)

A number of cost-saving strategies for reducing hospitalizations from nursing homes without compromising quality of care have been demonstrated in recent years. These include employing geriatric nurse practitioners in nursing homes,[33,34] providing financial incentives for primary care physicians and nursing home staff to treat serious acute illnesses on site,[35] offering nursing home residents the opportunity to express their preference with regard to use of acute hospital services in the event of a life-threatening acute illness,[28] and, in larger institutions, establishing an "infirmary" unit for care of serious acute illnesses.[36]

Most of these strategies require modification of traditional staffing and reimbursement for clinical care in nursing homes.[14,37,38] Much neglected in the past, these needs have begun to be met in the form of the 1989 legislation authorizing Medicare reimbursement of nurse practitioners in nursing homes (Public Law 101-239) and the 1993 implementation of increased Medicare fees for physician services in nursing homes.[39] Although these are important enabling steps, it is not likely that reimbursement reform alone, in the absence of organizational reform, will achieve the full potential for reducing transfers from nursing homes to hospitals. In this regard, the proposed inclusion of long-term care in national health care reform[40,41] affords a unique opportunity to overcome our long and costly legacy of separation of the acute and chronic health service sectors, epitomized by frequent transfers of patients between nursing homes and hospitals. Such integration of the delivery system in Great Britain has resulted in far lower rates of hospitalization from long-term care institutions.[15] This approach has been emulated in other societies with comprehensive national health programs[42] and deserves high priority when considering the needs of older Americans as part of national health care reform. □

TABLE 6—Distribution of Hospitalizations of Persons 65 Years of Age or Older from Nursing Homes and from the Community, by Principal Discharge Diagnosis

Principal Discharge Diagnosis	Nursing Home,[a] (n = 808) No.	%	Community,[b] (n = 24 891) No.	%
Cardiovascular*	126	15.6	5 548	22.3
Gastrointestinal tract	107	13.2	2 938	11.8
Pneumonia*	78	9.7	636	2.6
Hip fracture*	76	9.4	590	2.4
Genitourinary tract	59	7.4	1 430	5.7
Neoplasm*	40	5.0	3 119	12.5
Cataract surgery*	38	4.7	1 924	7.7
All other	284	35.1	16 185	35.0

[a]Hospitalizations for nursing home residents less than 65 years old are excluded.
[b]Monroe County, 1983.
*$P < .001$.

TABLE 7—Length of Stay, Charges, and Outcome of Hospitalizations of Persons 65 Years of Age or Older from Nursing Homes and the Community

Age Group, y	Nursing Home[a] No.	Average Length of Stay, d	Average Charge, $	Dead at Discharge, %	Community (1983)[b] No.	Average Length of Stay, d	Average Charge, $	Dead at Discharge, %
65–74	100	14.3	5812	8.0	12 586	11.5	5373	6.0
75–84	342	15.6	6312	13.7	9 046	13.8	5667	8.9
85+	366	13.3	5529	16.7	3 249	17.8	6133	14.1
All 65+	808	14.4	5849	14.4	24 891	13.2	5580	8.1

[a]This analysis does not include hospitalizations among nursing home residents less than 65 years old.
[b]Data on hospitalizations from the community were provided by INFOMED, Inc, an affiliate of the Rochester Area Hospital Corporation.

Acknowledgments

This research was supported by the National Institute on Aging (grant AG05423).

Preliminary data were presented at the 117th Annual Meeting of the American Public Health Association, October 22–26, 1989, Chicago, Ill.

We would like to acknowledge Ms Ann Treat, former executive director of the Genesee Valley Medical Foundation; Ms Lynn Varricchio, senior planner for chronic care of the Finger Lakes Health Systems Agency; and the medical records and billing offices of the community hospitals in Monroe County, New York, for facilitating data collection. We thank Dana Mukamel, PhD, for review and comment on the manuscript.

References

1. Kane RL, Mathias R, Sampson S. The risk of placement in a nursing home after acute hospitalization. *Med Care.* 1983;21:1055–1061.
2. Shaughnessy PW, Kramer AM. The increased needs of patients in nursing homes and patients receiving home health care. *N Engl J Med.* 1990;322:21–27.
3. Gillick M, Steel K. Referral from long-term care to acute facilities. *J Am Geriatr Soc.* 1983;31:74–78.
4. Irvine PW, Van Buren N, Crossley K. Causes for hospitalization for nursing home residents. The role of infection. *J Am Geriatr Soc.* 1984;32:103–107.
5. Gabow PA, Hutt DM, Baker S, et al. Comparison of hospitalization between nursing home and community residents. *J Am Geriatr Soc.* 1985;33:524–529.
6. Tresch DD, Simpson WM, Burton JR. Relationship of long-term and acute care facilities: the problem of patient transfer and continuity of care. *J Am Geriatr Soc.* 1985;33:819–826.
7. Lewis MA, Kane RL, Cretin S, et al. The immediate and subsequent outcomes of nursing home care. *Am J Public Health.* 1985;75:758–762.
8. Kayer-Jones JS, Weiner CL, Barbaccia JC. Factors contributing to the hospitalization of nursing home residents. *Gerontologist.* 1989;29:502–510.
9. *The National Nursing Home Survey: 1977.* Hyattsville, Md: National Center for Health Statistics; 1979.
10. National Center for Health Statistics. Discharges from nursing homes. Preliminary data from the 1985 National Nursing Home Survey. *Adv Data Vital Health Stat.* 1987;142. DHHS publication PHS 87-1250.
11. Jencks SF, Kan T. Do frail, disabled, and very old Medicare beneficiaries have higher hospital charges? *JAMA.* 1987;257:198–202.
12. Gillick MR, Serrell NA, Gillick LS. Adverse consequences of hospitalization in the elderly. *Soc Sci Med.* 1982;16:1033–1038.
13. Creditor MC. Hazards of hospitalization of the elderly. *Ann Intern Med.* 1993;118:219–223.
14. Vladek BC. *Unloving Care. The Nursing Home Tragedy.* New York, NY: Basic Books Inc; 1980.
15. Barker WH. *Adding Life to Years. Organized Geriatric Services in Great Britain and Implications for the United States.* Baltimore, Md: Johns Hopkins University Press; 1987.
16. Sirrocco A. An overview of the 1982 National Master Facility Inventory Survey of nursing and related care homes. *Adv Data Vital Health Stat.* September 20, 1985; no. 111. DHHS publication PHS 85-1250.
17. Hawes C. The Institute of Medicine Study: improving quality of nursing home care. In: Katz PR, Kane RL, Mezey MD, eds. *Advances in Long-Term Care.* Vol. 1. New York: Springer Publishing Co; 1991:147–168.
18. Eggert GM, Bowlyow JE, Nichols CS. Gaining control of the long-term care system: first returns from the ACCESS experiment. *Gerontologist.* 1980;20:356–363.
19. Foley WJ, Mengr MS, Schneider DP. A comparison of level of care predictions of six long-term care patient assessment systems. *Am J Public Health.* 1980;70:1152–1161.
20. Zimmer JG. Characteristics of patients and care provided in health-related and skilled nursing facilities. *Med Care.* 1975;13:992–1010.
21. Block JA, Regenstreif DI, Griner PF. A community hospital payment experiment outperforms national experience. The hospital experimental payment program in Rochester, N.Y. *JAMA.* 1987;257:193–197.
22. Rubenstein LZ, Ouslander JG, Wieland D. Dynamics and clinical implications of the nursing home-hospital interface. *Clin Geriatr Med.* 1988;4:471–491.
23. Bowlyow J. Acute and long-term care linkages. *Med Care Rev.* 1990;47:75–103.
24. Hall WJ, Griner PF. Cost-effective health care: the Rochester experience. *Health Aff.* Spring 1993:58–69.
25. Keeler EB, Kane RL, Solomon DH. Short and long-term residents of nursing homes. *Med Care.* 1981;19:363–369.
26. Uhlmann RF, Clark H, Perlman RA, et al. Medical management decisions in nursing home patients. Principles and policy recommendations. *Ann Intern Med.* 1987;106:879–885.
27. Danis M, Southerland LI, Garrett JM, et al. A prospective study of advance directives for life-sustaining care. *N Engl J Med.* 1991;324:882–888.
28. Mott PD, Barker WH. Hospital and medical care use by nursing home patients: the effect of patient care plans. *J Am Geriatr Soc.* 1988;36:47–53.
29. Kane RL, Olsen DM, Thetford C, Byrnes N. The use of utilization review records as a source of data on nursing home care. *Am J Public Health.* 1976;66:778–782.
30. Charleson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40:373–383.
31. Anderson GF, Steinberg EP. Predicting hospital readmissions in the Medicare population. *Inquiry.* 1985;22:251–258.
32. VanBuren CB, Barker WH, Zimmer JG, Williams TF. Acute hospitalization of nursing home patients: characteristics, costs and potential preventability. *Gerontologist.* 1982;22:129. Abstract.
33. Master RL, Feltin M, Jeinchill J, et al. A continuum of care for the inner city. Assessment of its benefits for Boston's elderly and high-risk populations. *N Engl J Med.* 1980;302:1434–1440.
34. Kane RL, Garrard J, Shay CL, et al. Effects of a geriatric nurse practitioner on process and outcome of nursing home care. *Am J Public Health.* 1989;79:1271–1277.
35. Zimmer JG, Eggert GM, Treat A, et al. Nursing homes as acute care providers: a pilot study of incentives to reduce hospitalizations. *J Am Geriatr Soc.* 1987;35:124–129.
36. Gordon M, Cheung M, Weisenthal S. An acute unit in a multilevel geriatric facility. *J Am Geriatr Soc.* 1990;38:728.
37. Rango N. Nursing home care in the United States. *N Engl J Med.* 1982;307:883–889.
38. Kayser-Jones JS. Physicians and the care of nursing home residents. In: Cassel CK, Walsh JR, eds. *Geriatric Medicine.* New York, NY: Springer-Verlag; 1984;2:397–412.
39. Tangalos EG, Stone D. The Medicare fee schedule in long-term care. *J Am Geriatr Soc.* 1993;41:574–575.
40. Harrington C, Cassel C, Estes CL, Woolhandler S, Himmelstein DU. A national long-term care program for the United States. *JAMA.* 1991;266:3023–3029.
41. Caring for the uninsured. Choices for reform. *JAMA.* 1991;265:2563–2565. Editorial.
42. Barker WH. Geriatrics internationally. In: Fox R, Puxty J, eds. *Medicine in the Frail Elderly.* London, England: Edward Arnold; 1993:241–248.

APPENDIX—Definitions of Skilled and Intermediate Care Facilities

Skilled Nursing Facility	Intermediate Care Facility
1. Residents with chronic medical conditions and disabilities requiring care of skilled technical/professional personnel	1. Residents with chronic medical or mental conditions requiring health-related institutional services but not skilled care
2. Physician visit at admission and at 30- or 60-day intervals	2. Physician visits at 60- or 90-day intervals
3. Licensed nursing coverage 24 hours a day, including registered nurse supervision during the day shift	3. Supervision by a registered nurse or licensed practical nurse during the day shift
4. Utilization review at regular intervals	4. Utilization review at regular intervals

Source. Adapted, with permission, from Appendix B, "Existing SNF Conditions of Participation and ICF Standards." In: Institute of Medicine. *Improving the Quality of Care in Nursing Homes.* Washington, DC: National Academy Press; 1986. © 1986 National Academy of Sciences.

Risk Factors for Delayed Immunization among Children in an HMO

ABSTRACT

Objectives. Improving the timely delivery of childhood immunizations has become a national imperative. This study aimed to identify nonfinancial predictors of delayed immunization among patients with good financial access to preventive care.

Methods. This prospective cohort study used telephone interviews and a computerized immunization tracking system to evaluate 13-month-old children (n = 530) in a regional group-model health maintenance organization.

Results. More than one third of parents interviewed did not know when the next immunization was due. Thirteen percent were late for the measles-mumps-rubella immunization, recommended at 15 months of age, by 90 days or more. Independent predictors of delayed immunization included having a larger number of children (odds ratio [OR] = 1.4, $P < .01$), not having a regular doctor (OR = 2.9, $P < .05$), not knowing when the shot was due (OR = 2.0, $P < .01$), and not worrying about the risks of shots (OR = 1.4, $P < .05$).

Conclusions. Financial access alone does not guarantee timely childhood immunization. In managed care settings, which may cover increasing numbers of children under health care reform, interventions are needed to better inform parents of when immunizations are due. (*Am J Public Health.* 1994;84:1621–1625)

Tracy A. Lieu, MD, MPH, Steven B. Black, MD, Paula Ray, MPH, Marie Chellino, MPH, Henry R. Shinefield, MD, and Nancy E. Adler, PhD

Introduction

More than one third of US children have not been appropriately immunized by the second birthday.[1] Delayed immunization is a major public health problem that has been associated with measles epidemics in recent years.[2] Better financial access to preventive care is a necessary but not sufficient condition for timely immunization.[3–6]

Improving immunization delivery will require a better understanding of nonfinancial reasons for delay. Previous studies have not been able to prospectively evaluate parent beliefs while controlling for financial access.[7–9] In this prospective study of children with good financial access to preventive care, we evaluated (1) the rate of delayed immunization; (2) the influences of nonfinancial risk factors; and (3) the accuracy of existing computerized information as a basis for identifying high-risk children for advance outreach efforts.

Method

Study Design

Parents of 13-month-olds were interviewed by telephone. Each child was then followed for at least 5 months via a computerized immunization tracking system. The system was initiated at the Northern California Kaiser Permanente Medical Care Program (KPMCP) in the winter of 1991 and included any child immunized at any of 13 clinics on-line at the start of this study. The measles-mumps-rubella (MMR) immunization was used as the main outcome measure because it is the first immunization due in the second year of life, when children are most likely to miss immunizations. The immunization was considered delayed if given more than 90 days after the due date.[10,11]

Study Population

KPMCP is a group-model health maintenance organization (HMO) that covers approximately 700 000 children in a geographic region approximately 250 miles in diameter. Most members receive health insurance as an employment benefit. Immunizations are covered in full, although some members must copay up to $15 for well-child care visits.

Between March and July of 1992, 6696 children seen in the KPMCP clinics served by the immunization tracking system turned 13 months old. Telephone interviews were attempted with 831 randomly selected families. Families were ineligible if the parents did not speak English, Spanish, or Cantonese (n = 10); the telephone number was wrong or had been disconnected (n = 52); they reported not using KPMCP, usually because they had a second form of insurance coverage (n = 61); or the child experienced a gap in KPMCP insurance coverage between 13 and 18 months of age (n = 103).

Tracy A. Lieu is with the Division of Research, The Permanente Medical Group, Oakland, Calif, and the Robert Wood Johnson Clinical Scholars Program and the Department of Pediatrics, University of California, San Francisco. Steven B. Black, Paula Ray, Marie Chellino, and Henry R. Shinefield are with the Kaiser Permanente Pediatric Vaccine Study Center, Oakland, Calif. Nancy E. Adler is with the Departments of Psychiatry and Pediatrics, University of California, San Francisco.

Requests for reprints should be sent to Tracy A. Lieu, MD, MPH, Division of Research, The Permanente Medical Group, 3505 Broadway, Oakland, CA 94611.

This paper was accepted May 10, 1994.

Editor's Note. See related editorial by Bernier (p 1567) in this issue.

TABLE 1—Demographic and Family Organization Characteristics of Families of 13-Month-Olds (n = 530) and Associations with Delayed Immunization: Northern California Kaiser Permanente Medical Care Program, 1992

	No.[a] (%)	Relative Risk of Delayed Immunization[b]	95% CI	P[c]
Demographics				
Race/ethnicity				
White	265 (50)	...	...	...
Latino	99 (19)	1.2	0.6, 2.4	NS
Asian	90 (17)	1.8	1.0, 3.2	NS
Black	68 (13)	2.1	1.2, 3.8	.03
Other or unspecified	8 (1)	...	...	...
Father's education				.10
Less than high school	39 (7)	1.9	0.8, 4.2	...
High school graduate	148 (28)	1.5	0.9, 2.8	...
Some college	156 (29)	1.4	0.8, 2.5	...
College graduate or more	186 (36)	...	...	...
Annual household income				NS
< $30 000	133 (27)	1.6	0.8, 3.2	...
$30 000–$50 000	206 (42)	1.9	1.0, 3.6	...
> $50 000	149 (30)	...	...	...
Both parents employed[d]	322 (61)	0.8	0.5, 1.2	NS
Family organization				
No. of children				.007
1	222 (42)	...	...	...
2	184 (35)	1.0	0.6, 1.9	...
≥3	124 (24)	2.2	1.3, 3.7	...
Single-parent family	50 (9)	0.8	0.3, 1.8	NS
Years of residence at current address				
< 1 y	91 (17)	1.3	0.8, 2.1	NS
1–3	200 (38)	1.1	0.7, 1.7	NS
> 3	238 (45)	...	...	...

Note. CI = confidence interval; NS = not significant.
[a]Numbers may not sum to 530 owing to missing values. No more than 3% of data in any category were missing, except for income, in which 8% of data were missing.
[b]Relative risk is expressed in comparison with the baseline categories: White, college graduate or more, more than $50 000 annual household income, one child, one parent employed, two-parent family, and more than 3 years' residence at current address.
[c]The chi-square test was used for categorical variables. The Wilcoxon rank-sum test was used for ordinal variables (father's education, annual household income, and number of children in family).
[d]For one-parent households, this variable was scored "yes" if the parent was employed.

Telephone Survey

A 15-minute, 33-question telephone survey was conducted in English, Spanish, or Cantonese by six experienced interviewers. It covered demographic characteristics, family organization, health care system variables, and vaccine knowledge and attitudes. Family organization questions included numbers of children and adults in the household and how many years the family had lived at its present address. Health care system questions covered experiences at the last visit for well-child care and whether the child had a regular doctor or nurse-practitioner at KPMCP.

For vaccine knowledge and attitudes, parents were asked when they expected the child's next well-child checkup, when they expected the next shot, whether the child or any of the family's other children had ever had a worrisome reaction to a shot, and whether the child had ever been more than a month late for a shot. Parents were asked how much they worried about the risks of shots, how much they would worry if the child were not up to date on shots, and whether the parent had ever hesitated to have the child get a shot. (A copy of the survey instrument is available from the authors.)

A follow-up survey was conducted for children without an immunization tracking system record of an MMR by 18 months of age. For 13 of these children, the parent reported and chart review confirmed that the MMR immunization had been given, usually at a KPMCP facility that was not yet part of the immunization tracking system. The computerized data were corrected prior to analysis.

Statistical Analysis

In the bivariate analysis the chi-square test (with Yates' continuity correction for 2×2 tables) was used for categorical variables, the Wilcoxon rank-sum test was used for ordinal variables, and the Spearman correlation coefficient was used for associations between ordinal variables. Multivariate analysis was performed with logistic regression procedures in SAS (SAS Institute, Inc, Cary, NC).

Results

Of the 605 eligible families, 530 (88%) completed interviews. Sixty-eight were unreachable after repeated attempts, and 7 declined to be interviewed.

Predictor Variables

Table 1 shows the demographic and family organization characteristics of the study population. In terms of health care system variables, almost all parents (96%) said their child had a regular primary care provider at KPMCP. Most (53%) had no copayment for well-child visits; 43% paid between $1 and $5. Forty-five percent of parents had missed or rescheduled work to bring their child to his or her last well-child visit. A plurality of parents (45%) had had the last well-child visit scheduled between 1 and 2 months in advance, 51% had waited less than 15 minutes to see the primary care provider, and 59% had spent between 1 and 2 hours on the round trip to the appointment, including transit time.

Table 2 shows parental knowledge and attitudes about immunization. Parents were asked when they expected their child's next well-child appointment and next shot. Answers that fell before the child's 16-month birthday were scored as correct. It is noteworthy that 36% of the parents in this study did not know when their child's next shot was due; this answer was highly correlated with not knowing when the next well-child visit was due ($P < .001$).

Delayed Immunization Rates and Predictors of Delay

Two thirds of the children in the study group received their MMR vaccinations within 30 days of the due date. However, 16% were late for the vaccination by 60 days or more, and 13% were late by 90 days or more. The rate of delayed immunization at 90 days was higher among ineligible and unreachable families (21%) than among the study group.

The baseline telephone interview was not intended to change behavior, but it could have influenced study group families to seek more timely immunization. After children with gaps in KPMCP insurance between 13 and 18 months of age (for whom the immunization tracking system might have incomplete records) were excluded, there were 530 families in the study group, 645 in the attempted interview group, and 4448 in the control group not selected for interview attempts. The study group had a lower rate of delayed immunization (13%) than the control group (19%; $P < .001$), but this comparison is biased by the inability to exclude ineligible and unreachable patients from the control group. The rate of delayed immunization among all 645 families in the attempted interview group (16%) was lower but not significantly different from that of the control group ($P = .08$).

Univariate analysis of predictors. Black race was correlated with delayed immunization (Table 1): 21% of Black children, compared with 10% of White children, were late by 90 days or more ($P < .05$). Income, education, and language spoken at home were not significantly associated with immunization delay. Families with more children were at higher risk for delayed immunization ($P < .01$).

Unexpectedly, parents who reported having had the previous appointment scheduled more than 2 months in advance had lower rates of delayed immunization than parents who had the appointment scheduled sooner ($P < .05$). Having had the previous appointment scheduled farther in advance may indicate better organizational skills on the part of parents rather than a longer wait for appointments. The amount of copayment was not associated with delayed immunization.

Most attitudes toward vaccines were not associated with delayed immunization (Table 2). Paradoxically, the more the parent worried about the risks of vaccines, the lower the likelihood of delayed immunization ($P < .05$). Not knowing when the next shot was due ($P < .001$) was strongly correlated with delayed immunization. It should be noted that 15-month-olds are eligible to receive the MMR immunization on time even if they have been late for previous immunizations. Despite this fact, having been late for a previous shot by the parent's report was also a strong predictor of delayed MMR immunization ($P < .001$).

Multivariate analysis of predictors. Variables associated with delayed immunization at a level of $P < .10$ were included in a logistic regression model (Table 3). Families with more children, those who did not identify a regular doctor, those who did not know when the shot was due, and those who worried less about the risks of shots were at significantly increased risk for delayed immunization. After adjusting for these other factors, Black race was no longer a significant predictor of delayed immunization. Having been late for a previous immunization was excluded from the final model because its similarity to the outcome of interest tended to reduce the explanatory power of other predictors. However, when this variable was included it was significant (adjusted odds ratio = 3.0, 95% confidence interval = 1.6, 5.7, $P < .001$), and the other significant predictors remained significant with the exception of not identifying a regular doctor ($P = .07$).

TABLE 2—Parental Knowledge and Attitudes and Their Associations with Delayed Immunization of 13-Month-Olds (n = 530): Northern California Kaiser Permanente Medical Care Program, 1992

	No. (%)	Relative Risk of Delayed Immunization[a]	95% CI	P[b]
Late for a previous shot by a month or more (by parent's report)				
Yes	99 (19)	2.6	1.7, 4.1	<.001
No	421 (80)	...	...	...
Don't know	10 (2)	...	...	...
Reason for being late for previous shot				
Child was sick	31 (31)	...	...	...
Appointment was unavailable	13 (13)	...	...	...
Too busy	7 (7)	...	...	...
Didn't know it was due	3 (3)	...	...	...
Forgot	3 (3)	...	...	...
Other[c]	42 (42)	...	...	...
Did not know when next shot was due	189 (36)	2.2	1.4, 3.3	<.001
Did not know when next well-child visit was due	149 (28)	1.6	1.0, 2.5	.06
Ever hesitated to have a previous shot	33 (6)	1.7	0.9, 3.5	NS
Ever had a vaccine reaction that caused worry	60 (11)	0.8	0.3, 1.7	NS
Amount parent worries about risks of shots				.02
A great deal	35 (7)	...	...	...
A moderate amount	111 (21)	1.0	0.3, 3.6	...
A little	215 (41)	1.4	0.4, 4.3	...
Not at all	169 (32)	2.0	0.7, 6.4	...
Amount parent would worry if child were late for a vaccine				NS
A great deal	314 (60)	1.5	0.8, 2.8	...
A moderate amount	126 (24)	...	...	...
A little	65 (12)	2.1	1.0, 4.5	...
Not at all	23 (4)	2.0	0.7, 5.7	...

Note. CI = confidence interval; NS = not significant.

[a]Relative risk is expressed in relation to the baseline categories: not late for a previous shot, knew when next shot or well-child visit was due, never hesitated to have a previous shot, never had a vaccine reaction that caused worry, worries a great deal about risks of shots, and would worry a moderate amount if child were late for a vaccine.

[b]The chi-square test was used for categorical variables. The Wilcoxon rank-sum test was used for ordinal variables (amount parent worries about risks of shots, amount parent would worry if child were late for a vaccine).

[c]Other reasons parents gave for being late for shots included family emergencies, lapses in health insurance coverage, the regular doctor's going on vacation, and rescheduling of appointments.

TABLE 3—Logistic Regression Analysis of Predictors of a Delay of 90 Days or More in Measles-Mumps-Rubella Immunization: Northern California Kaiser Permanente Medical Care Program, 1992

	Adjusted Odds Ratio	95% Confidence Interval	P
Black race	2.0	0.9, 4.3	NS
Asian race	1.6	0.8, 3.3	NS
Latino race/ethnicity	1.1	0.5, 3.4	NS
Number of children in family	1.4	1.1, 1.8	.007
How far in advance previous appointment was made[a]	0.7	0.5, 1.0	NS
No regular doctor by parent's report	2.9	1.0, 8.6	.05
Didn't know when next shot was due	2.0	1.2, 3.5	.01
Amount parent worries about risks of shots[b]	0.7	0.6, 0.8	.03

[a]Ordinal variable. 1 = less than 1 month, 2 = 1 to 2 months, 3 = more than 2 months. Thus, having made the last appointment farther in advance was associated with a lower risk of delayed immunization.
[b]Ordinal variable. 1 = not at all, 2 = a little, 3 = a moderate amount, 4 = a great deal. Thus, worrying more about the risks of vaccines was associated with lower rates of delayed immunization.

Risk Prediction Based on Computerized Data

Two variables available from traditional computerized databases—not having a regular physician assigned and not having made at least three well-child visits during the first 7 months of life—were used as predictors of delayed MMR immunization. Having either of these risk factors identified children who had delayed immunization with a sensitivity of 41% and a specificity of 75%. Thus, fewer than half of the children who would be late for the MMR immunization could be identified in advance by means of information from existing databases.

Follow-Up Survey

Of the 69 families identified by the immunization tracking system as 90 or more days late for the MMR, 43 had follow-up interviews. Two children were reportedly immunized outside KPMCP, one in Mexico and one at a public clinic. Nine parents said a shot had been given at KPMCP, but there was no record of these shots in the written chart or on the tracking system. Three children had received the MMR after 18 months of age.

Twenty-nine parents said their child had not received a vaccination in the past 3 months or that they did not know. Sixteen of these said they had been aware the shot was due and 12 said they had not. The main reason cited for delay was not knowing the shot was due (n = 10), illness of the child (n = 2), and being too busy or forgetting (n = 4). Eight parents cited other reasons, including difficulty in taking time off from work and expecting reminders in the mail.

Discussion

This study shows that financial access to preventive care does not guarantee timely childhood immunization. Although prompt immunization rates are higher among KPMCP children than among the general population, one in eight children enrolled in KPMCP was at least 3 months late for the MMR vaccination.

This study suggests that managed care settings should attempt to better inform parents of when immunizations are due. More than one third of parents of 13-month-olds in this study did not know when their child's next shot was due, and these families were twice as likely as others to be late for the MMR. These findings support two interventions: (1) improving the written information handed out to parents at visits to highlight when immunizations are due and (2) reminding families when immunizations are due via letters or telephone calls.

Advance outreach efforts in managed care settings may be difficult; variables from existing computerized databases could not accurately predict which children would have delayed immunization in this study. However, immunization tracking systems should at least enable providers to promptly remind families who have already missed an immunization. An automated telephone reminder system has proven cost-effective when used with an immunization tracking system in the general population.[12]

This study underscores the importance of continuity of health care, in that families who did not identify a regular doctor or nurse-practitioner were almost three times as likely to be late for the MMR immunization. This result supports evidence from other studies, which suggests that the ability of patients to identify a primary care provider enhances the quality of health care.[13-16] The study did not find an association between copayment amounts and delayed immunization; previous studies have reached divided conclusions on this question.[17,18]

The relatively high immunization rates in this population may be explained, at least in part, by good financial access to a comprehensive health care system that covered immunizations. This study is unique in controlling for financial access while prospectively evaluating nonfinancial predictors. Education, income, and language were not associated with delayed immunization. Families with more children were more likely to have delayed immunization; this finding was similar to that of a previous study.[8] In the general population, children of minority races are at higher risk for delayed immunization.[19] In the present study, Black race was associated with delayed immunization, but the association was not statistically significant after other potentially mediating variables, such as knowledge of when the shot was due, ability to identify a regular doctor, and family size, were controlled.

We found no association between most parental attitudes toward vaccination and subsequent delay in immunization. Unexpectedly, parents who worried more about the risks of shots were actually less likely than others to be late, perhaps because their worrying reflected a generalized anxiety about health that could be associated with a higher likelihood of seeking preventive services.

This study included families of diverse races and educational levels. However, it was limited by the ineligibility of 27% and the unreachability of 9% of the patients selected for attempted interviews. Hard-to-contact respondents may have been more likely to be non-White[20] or to have recently moved or changed insurance plans. This could have biased the study against identifying these characteristics as predictors of delayed immunization. The finding that 24% of the children experienced gaps in KPMCP insurance between 13 and 18 months of age demonstrates the challenge of devel-

oping integrated immunization tracking systems, which will need to coordinate records as families change providers and insurance plans.

We conclude that delays in childhood immunization and gaps in parent understanding of immunization requirements occur even among families with good financial access to health care. Financing reform should improve access to immunizations for some children. Among children with good financial coverage, more work is needed on interventions to better inform parents of when immunizations are due. □

Acknowledgments

This study was supported in part by a grant from the Centers for Disease Control and Prevention. Dr Lieu was supported by the Robert Wood Johnson Foundation Clinical Scholars Program.

We are grateful to Melanie Gould, Judy Vasos, Athena Burns, Socorro Ramirez, and Marie Wong, who conducted the interviews for this study. We appreciate the statistical advice of Bruce Fireman, MA, and the programming assistance of Edwin M. Lewis, MPH, and Joan Schwalbe, MPH. We thank Lyn Wender for editing and Robert Hiatt, MD, PhD, and the Robert Wood Johnson Clinical Scholars of the University of California, San Francisco, for their advice on the study design and on the manuscript.

References

1. Williams BC. Immunization coverage among preschool children: the United States and selected European countries. *Pediatrics*. 1990;86:1052–1055.
2. National Vaccine Advisory Committee. The measles epidemic. *JAMA*. 1991;266:1547–1552.
3. Freed GL, Katz SL. The comprehensive childhood immunization act of 1993: toward a more rational approach. *N Engl J Med*. 1993;329:1957–1960.
4. Lopreiato JO, Ottolini MC. Does universal access to care guarantee immunization compliance? *Am J Dis Child*. 1993;147:447. Abstract.
5. Marwick C. Improving preschool immunization levels proving more complex than simply providing vaccines. *JAMA*. 1993;269:1480–1481.
6. Fielding JE, Cumberland WG, Pettitt L. Immunization status of children of employees in a large corporation. *JAMA*. 1994;271:525–530.
7. Marks JS, Halpin TJ, Irvin JJ, Johnson DA, Keller JR. Risk factors associated with failure to receive vaccinations. *Pediatrics*. 1979;64:304–309.
8. Bobo JK, Gale JL, Thapa PB, Wassilak SGF. Risk factors for delayed immunization in a random sample of 1163 children from Oregon and Washington. *Pediatrics*. 1993;91:308–314.
9. Rodewald LE, Szilagyi PG, Humiston SG, et al. Factors associated with undervaccination of preschool-age children: a case–control study. *Pediatr Res*. 1993;33:122A. Abstract.
10. Dietz VJ, Zell ER, Stevenson J. Defining delayed immunization. *Pediatr Infect Dis J*. 1993;12:353–354. Letter.
11. Varasso DA, Redlener I. Defining delayed immunization. *Pediatr Infect Dis J*. 1992;897. Letter.
12. Dini E. Informing parents using registries and reminders: the Georgia Autodialer Study. Presented at the Institute of Medicine Workshop on Overcoming Barriers to Immunization; December 8–9, 1993; Washington, DC.
13. Alpert JJ, Robertson LS, Kosa J, Heagarty MC, Haggerty RJ. Delivery of health care for children: report of an experiment. *Pediatrics*. 1976;57:917–930.
14. Becker MH, Drachman RH, Kirscht JP. Predicting mothers' compliance with pediatric medical regimens. *J Pediatr*. 1972;81:843–854.
15. Charney E, Bynum R, Eldredge D, et al. How well do patients take oral penicillin? A collaborative study in private practice. *Pediatrics*. 1967;40:188–195.
16. Wasson JH, Sauvigne AE, Mogielnicki P, et al. Continuity of outpatient medical care in elderly men. *JAMA*. 1984;252:2413–2417.
17. Cherkin DC, Grothaus L, Wagner EH. The effect of office visit copayments on preventive care services in an HMO. *Inquiry*. 1990;27:24–38.
18. Lurie N, Manning WG, Peterson C, Goldberg GA, Phelps CA, Lillard L. Preventive care: do we practice what we preach? *Am J Public Health*. 1987;77:801–804.
19. Orenstein WA, Atkinson W, Mason D, Bernier RH. Barriers to vaccinating preschool children. *J Health Care Poor Underserved*. 1990;1:315–330.
20. Mishra SI, Dooley D, Catalano R, Serxner S. Telephone health surveys: potential bias from noncompletion. *Am J Public Health*. 1993;83:94–99.

Ninth National Conference on Chronic Disease Prevention and Control

The Centers for Disease Control and Prevention, Association of State and Territorial Health Officials, and Association of State and Territorial Chronic Disease Program Directors will cosponsor the Ninth National Conference on Chronic Disease Prevention and Control, which will be held December 7–9, 1994, in Washington, DC. The conference is open to the public. There is no registration fee. This year's theme is "Chronic Disease Prevention and Control and the New Public Health."

The conference will provide participants with a forum for sharing information, skills, knowledge, and experiences related to chronic disease prevention and control. The conference will emphasize interactions among federal, state, and local health departments; voluntary health agencies; and professional organizations. Presentations will be given in plenary, invited, abstract, poster, training, and workshop sessions.

Additional information is available from the National Center for Chronic Disease Prevention and Health Promotion, CDC, Mailstop K-30, 4770 Buford Hwy NE, Atlanta, GA 30341-3724; tel (404) 488-5390; fax (404) 488-5974.

Presumptive Eligibility for Pregnant Medicaid Enrollees: Its Effects on Prenatal Care and Perinatal Outcome

Joyce M. Piper, DrPH, Edward F. Mitchel, Jr, MS, and Wayne A. Ray, PhD

ABSTRACT

Objectives. "Presumptive eligibility" permits pregnant prospective Medicaid enrollees to obtain services during the application period. The purpose of this study was to assess the effects of presumptive eligibility on the receipt of prenatal care and the occurrence of low-birthweight births and neonatal, perinatal, and infant mortality.

Methods. Outcome rates for pregnant women who enrolled in Tennessee Medicaid in the 6-month period before presumptive eligibility was enacted were compared with those obtained for pregnant women who enrolled in the 6-month period after presumptive eligibility had been in effect for 5 months.

Results. Women in the "after" group were 40% more likely to enroll and 30% more likely to obtain prenatal care in the first trimester. They were 300% more likely to fill a prescription for prenatal vitamins in the first trimester and 16% more likely to have begun prenatal care before the third trimester. However, they were similar to those enrolling in the "before" time period in terms of the occurrence of adverse perinatal outcomes.

Conclusions. When barriers to prenatal care, including bureaucratic ones, are removed, low-income women will seek care earlier and more frequently. (*Am J Public Health.* 1994;84: 1626–1630)

Introduction

Expanded Medicaid eligibility for pregnant women is based on the premise that early, regular prenatal care reduces adverse pregnancy outcomes.[1] Low-income women, who are more likely to start pregnancy with unmet health care needs, have been shown to derive particular benefit from prenatal care.[2,3] Counseling about diet, smoking cessation, and drug avoidance is an important component of prenatal care and its impact may be greatest if counseling is received early.[4] Early prenatal vitamin supplementation may also be particularly important for low-income women.[5]

"Presumptive eligibility" is a term used in federal legislation permitting states to allow certain providers (such as local health departments, community health centers, and hospitals) to determine temporary eligibility and to provide prenatal care during the formal processing period.[6] In Tennessee, selected providers may issue temporary (45-day) Medicaid cards to pregnant clients who appear to qualify for Medicaid. These cards can then be used to obtain outpatient services, including prenatal care, while the permanent Medicaid enrollment application is being processed.

The same federal legislation that permitted states to issue presumptive eligibility also authorized states to drop the "assets test" for pregnant clients, thus further allowing states to simplify the complex and lengthy application process and accompanying documentation. Prior to these legislative changes, a pregnant prospective Medicaid enrollee would be required to apply for enrollment, supply detailed assets information, and have her application evaluated. Only after completion of this process, which often required several visits to the application office, could the applicant schedule a prenatal visit with a health care provider or clinic willing to accept Medicaid patients. As a result, it might be impossible for her to use Medicaid to obtain early (first trimester) prenatal care.

As part of a series of investigations designed to assess the effects on pregnancy outcomes of Tennessee Medicaid expansions for pregnant women,[7,8] we examined the effects of presumptive eligibility. Initially authorized by the Omnibus Budget Reconciliation Act of 1986, both presumptive eligibility and elimination of the assets test are options chosen by a majority of states.[9] Some states have opted to expedite Medicaid enrollment for pregnant women in other ways, such as permitting mail-in applications and shortening the form length.[9]

In Tennessee, Medicaid coverage of pregnant women was expanded to include those living at 100% of the federal poverty level on July 1, 1987; the assets test was eliminated on March 1, 1988; and presumptive eligibility went into effect on February 1, 1989. At that time, more than half the states did not have presumptive eligibility.[10] Therefore, our findings regarding presumptive eligibility and pregnancy outcome may predict subsequent changes in other states and the country as a whole.

At the time of the study, all authors were with the Preventive Medicine Department, Vanderbilt University, Nashville, Tenn. Dr Piper died after the study was completed.

Requests for reprints should be sent to Wayne A. Ray, PhD, Preventive Medicine Department, MCN-A1112, Vanderbilt University, Nashville, TN 37232-2637.

This paper was accepted March 2, 1994.

Note. The opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding agency.

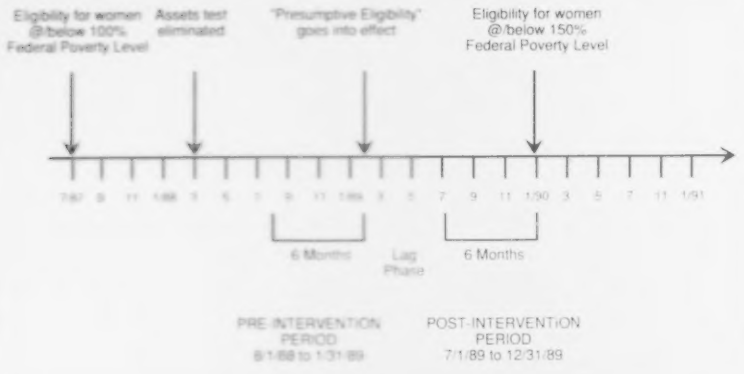

Eligibility for women @/below 100% Federal Poverty Level
Assets test eliminated
"Presumptive Eligibility" goes into effect
Eligibility for women @/below 150% Federal Poverty Level
6 Months
Lag Phase
6 Months
PRE-INTERVENTION PERIOD 8/1/88 to 1/31/89
POST-INTERVENTION PERIOD 7/1/89 to 12/31/89

FIGURE 1—Time line showing "before" ("preintervention") and "after" ("postintervention") study time periods in relation to Tennessee Medicaid regulatory changes for pregnant women.

Methods

Study Population

We studied outcome rates for pregnant women who enrolled in Tennessee Medicaid in the 6-month period before presumptive eligibility was enacted (August 1, 1988, through January 31, 1989) and for pregnant women who enrolled in Medicaid in the 6-month period after presumptive eligibility had been in effect for 5 months (July 1, 1989, through December 31, 1989). As shown in Figure 1, during both of these periods Tennessee Medicaid eligibility was restricted to those with countable income (no assets test) at or below 100% of the federal poverty level.

Because our purpose was to evaluate a regulatory change that affected only pregnant women who were not enrolled in Medicaid, we removed any woman from the study population who was already enrolled in Medicaid before the start of the index pregnancy. These high-risk women included women with chronic medical conditions, very young women who qualified for Medicaid as children themselves, women who became pregnant during the postpartum coverage period of a previous pregnancy, and very poor women who received Aid to Families with Dependent Children. The study population was further limited to women who were either White or Black and who had delivered a single live infant with a recorded birthweight of 500 to 6000 g or a stillborn infant.

Data Sources

Data were obtained from Tennessee birth, fetal death, and death certificates and Tennessee Medicaid files. Data available from birth and fetal death certificates included personal identifiers of the mother and infant (or fetus); demographic characteristics, including maternal age, race, marital status, and place of residence and number of years of education of both parents; and medical characteristics of the mother and infant (or fetus), including month of pregnancy prenatal care began, date of the last menstrual period, mother's parity, plurality of the birth, and birthweight. Data pertaining to demographics, the timing of initiation of care, and the adequacy of care were taken from birth and fetal death certificates. Approximately 1% of birth and fetal death certificates from both periods studied were excluded because of missing data. Death certificate information included personal identifiers, demographic characteristics of the decedent, and cause of death. Linked birth and death certificate files were used to ascertain mortality rates. Medicaid data files included personal identifiers, demographic characteristics, beginning and ending dates of Medicaid coverage, hospital and outpatient diagnoses and procedures, and outpatient prescription medications, including when the prescription was filled.[11] During the study period, prenatal vitamins were included in the Medicaid formulary. Data on the use of prenatal vitamins were taken from Medicaid files, as were data related to the timing of enrollment.

TABLE 1—Characteristics (%) of Pregnant Women Who Enrolled in Tennessee Medicaid before and after Enactment of Presumptive Eligibility

	"After" Group (n = 9702)	"Before" Group (n = 9495)	*P*[a]
Age group, y			.02
<18	13.2	13.9	
18–19	22.7	22.1	
20–24	37.4	38.6	
25–29	17.0	17.0	
≥30	9.7	8.5	
Race			.06
White	71.6	70.3	
Black	28.4	29.7	
Marital status			.90
Married	49.1	49.2	
Unmarried	50.9	50.8	
Education, y			.97
<12	44.0	44.3	
12	43.9	43.6	
13–15	10.6	10.6	
≥16	1.5	1.5	
Parity density			.80
First	48.5	48.4	
Low	14.8	14.5	
High	36.7	37.1	
Residence county			.82
Urban	64.6	64.8	
Rural	35.4	35.2	

[a]By chi-square test.

Linkage of Medicaid and Vital Records

We identified births to women in the study population through the computerized linkage of Medicaid, birth certificate, infant death certificate, and fetal death certificate files.[12] The linkage between the Medicaid files and vital records was required to ensure complete ascertainment of births among Medicaid-enrolled women and to obtain information on the women and their offspring that was available only from vital records. Because we did not require that the child be enrolled in Medicaid to classify the mother as enrolled at the time of birth, we included in the Medicaid group those women whose deliveries resulted in stillbirth or whose newborns died before they could be enrolled in Medicaid.[12] Consideration of only those births for which there is Medicaid information on the mother, delivery, and child may lead to serious

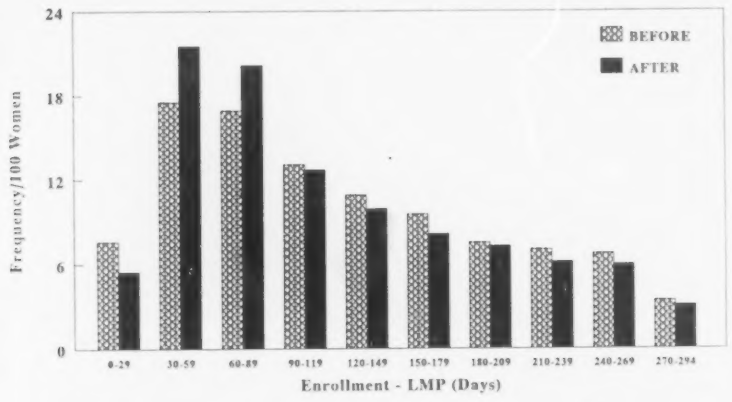


FIGURE 2—Time in pregnancy (days since last menstrual period [LMP]) Medicaid enrollment began for women who enrolled in Tennessee Medicaid in the study periods before and after presumptive eligibility was enacted.

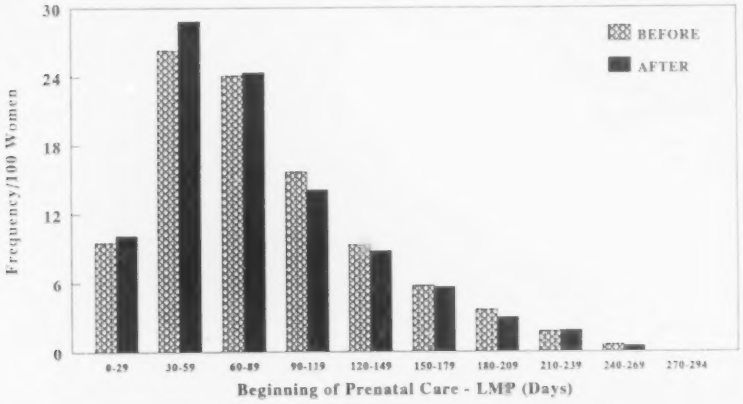


FIGURE 3—Time in pregnancy (days since last menstrual period [LMP]) prenatal care began for women who enrolled in Tennessee Medicaid in the study periods before and after presumptive eligibility was enacted.

underascertainment of fetal, perinatal, and neonatal mortality rates.[12]

Definitions

The start of pregnancy was defined as the date of the last menstrual period as reported on the birth or fetal death certificate. If the day of the month of the last menstrual period was missing, the 15th was imputed. For approximately 15% of linked deliveries, the last menstrual period either was missing month and/or year or, for live births, resulted in implausible gestational lengths (<140 days or >294 days). In these cases, the date of last menstrual period was estimated by using the median length of gestation for infants of similar race and birthweight. Reportable fetal deaths for which the date of the last menstrual period was missing were not studied. Reportable fetal deaths are fetal deaths for which a Tennessee Fetal Death Certificate must be filed. They include deaths of fetuses weighing 500 g or more or, in the absence of weight, of 20 or more weeks' completed gestation. All Tennessee counties were classified as either urban (within a metropolitan statistical area) or rural (outside metropolitan statistical areas) as designated by the Bureau of the Census in cooperation with the Tennessee State Planning Office.[13] Parity was defined as the total number of live births and fetal deaths, including the current birth. High parity density was defined as the third or later birth to women younger than 25 years and the fourth or later birth to women aged 25 through 29 years. All other births except the first represent low parity density. The first trimester was defined as the 90-day period after the date of the last menstrual period as reported on the birth certificate. Thus, the third trimester represents the 90-day period beginning 180 days after the date of the last menstrual period. The use of prenatal care is reported in terms of trimester of pregnancy in which prenatal care began (no first-trimester care, no care, or care begun in the third trimester), and inadequate care is measured by the Kessner Index.[14] However, birth certificate data may not be sufficiently accurate to rely on this or any of the indices of the adequacy of prenatal care.[15] Birthweights of less than 2500 g were classified as very low (<1500 g) or moderately low (1500–2499 g). Perinatal mortality rate was defined as the number of reportable fetal deaths plus the number of neonatal deaths divided by the total number of live births and reportable fetal deaths. Prenatal vitamins were defined as any vitamin or mineral product included in the Medicaid formulary.

Statistical Analyses

The chi-square test was used to test the null hypothesis (i.e., that there was no difference in the characteristics of Medicaid enrollees in the two time periods). The *t* test was used to test the hypothesis that presumptive eligibility shortens the time from initial Medicaid application to receipt of prenatal care. Three different aspects of timing were examined: (1) when Medicaid enrollment began, relative to the last menstrual period; (2) when prenatal care began, relative to the last menstrual period; and (3) when the woman initially filled a prescription for prenatal vitamins, relative to the last menstrual period.

To more fully assess the effects of individual and combined maternal factors in the comparison groups, we calculated odds ratios for measures of the attainment of prenatal care and adverse pregnancy outcome from the unconditional regression model obtained from the BMDP logistic regression program.[16] Models were constructed with terms that

included the study period and the maternal characteristics of age group, race, marital status, educational attainment, parity density, and urban/rural county of residence.

Results

In the 6-month "before" time period (August 1, 1988, through January 31, 1989), there were 9495 women who enrolled in Medicaid after their last menstrual period who otherwise met our study criteria. In the 6-month "after" time period (July 1, 1989, through December 31, 1989), there were 9702 women who enrolled in Medicaid after their last menstrual period who otherwise met our study criteria. Of these, 3995 women (41.2%) enrolled in Medicaid with presumptive eligibility.

Study subjects in the "before" and "after" time periods were similar in terms of race, marital status, educational level, parity, and county of residence (Table 1). However, slightly more women older than 29 years enrolled in the "after" than in the "before" time period (9.7% vs 8.5%; $P = .02$).

Figure 2 compares the distributions of the time in pregnancy when Medicaid enrollment began for the two study groups. Pregnant women who enrolled in Medicaid in the "after" time period (mean = 3.40 months, SE = 2.45) did so earlier in pregnancy than those in the "before" time period (mean = 3.59 months, SE = 2.53) ($t = 5.27, P = .00002$).

Figure 3 compares the distributions of the time in pregnancy when prenatal care began for the two study groups. When those who received no care (3.25% in the "after" group and 3.54% in the "before" group) were excluded, women in the "after" group began prenatal care slightly earlier (mean = 3.25 months, SE = 1.68) than those in the "before" group (mean = 3.34 months, SE = 1.70) ($t = 4.16, P = .00002$).

Figure 4 compares the distributions of the time in pregnancy when the woman initially filled a prescription for prenatal vitamins for the two study groups. When those who did not fill a prescription for prenatal vitamins at any time in the index pregnancy (45.46% in the "after" group and 54.60% in the "before" group) were excluded, women in the "after" group filled prescriptions for prenatal vitamins earlier in gestation (mean = 5.41 months, SE = 2.03) than women in the "before" group (mean = 6.09 months, SE = 1.86) ($t = 16.80, P = .00009$).

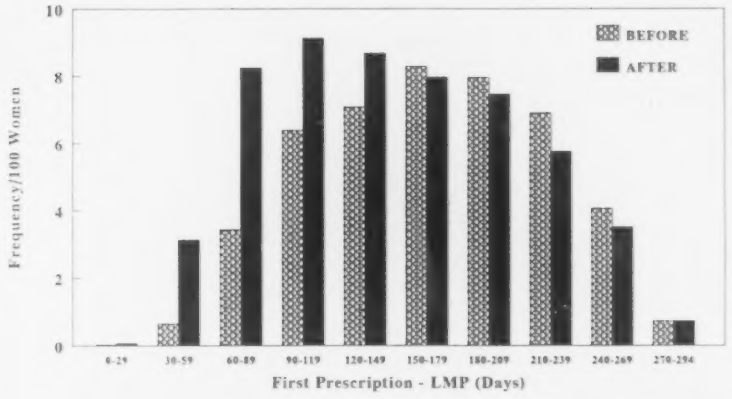


FIGURE 4—Time in pregnancy (days since last menstrual period [LMP]) initial prescription for prenatal vitamins and minerals was filled for women who enrolled in Tennessee Medicaid in the study periods before and after presumptive eligibility was enacted.

TABLE 2—Rates of Prenatal Care Use (%) for Pregnant Women Who Enrolled in Tennessee Medicaid before and after Enactment of Presumptive Eligibility

	"After" Group	"Before" Group	Adjusted Odds Ratio[a]	95% Confidence Interval
Enrolled in Medicaid in first trimester	50.2	42.0	1.40	1.32, 1.48
Prenatal care began in first trimester	64.2	60.9	1.30	1.17, 1.45
Prenatal vitamins in first trimester	11.4	4.1	3.06	2.71, 3.46
Had adequate or intermediate prenatal care	87.4	85.9	1.14	1.04, 1.24
Prenatal care began before third trimester	93.0	92.0	1.16	1.03, 1.29

Note. The variables in the model include maternal age group, race, education level, marital status, parity density, and county of residence.
[a]Reference category: the "before" time period.

TABLE 3—Rates (per 1000) of Adverse Pregnancy Outcomes for Pregnant Women Who Enrolled in Tennessee Medicaid before and after Enactment of Presumptive Eligibility

	"After" Group	"Before" Group	Adjusted Odds Ratio[a]	95% Confidence Interval
Moderately low birthweight (1500–2499 g)	70.1	75.4	0.91	0.81, 1.01
Very low birthweight (< 1500 g)	14.0	16.1	0.86	0.68, 1.09
Neonatal mortality	4.2	4.7	0.91	0.60, 1.40
Perinatal mortality	8.9	7.9	1.15	0.84, 1.57
Infant mortality	9.5	8.9	1.06	0.79, 1.43

Note. The variables in the model include maternal age group, race, education level, marital status, parity density, and county of residence.
[a]Reference category: the "before" time period.

Table 2 presents rates of prenatal care and adjusted odds ratios for the two study groups. Women in the "after" group were 40% more likely to enroll in the first trimester and 30% more likely to obtain prenatal care in the first trimester. They were more than three times as likely to fill a prescription for prenatal vitamins in the first trimester than were women in the "before" group. They were also more likely to have adequate or intermediate care as measured by the Kessner Index and to have begun prenatal care before the third trimester of pregnancy. However, there was no difference in adverse outcomes of pregnancy between the two groups (Table 3).

Discussion

We compared two groups of pregnant women: those who enrolled in Medicaid at a time when eligibility was limited to those whose countable income was at or below 100% of the federal poverty level and those who enrolled in Medicaid at a time when eligibility was similarly limited but enrollment was facilitated by presumptive eligibility. Presumptive eligibility increased the proportion of women who enrolled in Medicaid during the first trimester and was successful in reducing the proportion of women who failed to begin prenatal care in the first trimester, the proportion who obtained inadequate care (too few visits for gestational age at birth), and the proportion who delayed obtaining care until the third trimester or failed to get any care at all. It also led to a dramatic increase in first-trimester Medicaid prescriptions for prenatal vitamins, although this finding must be tempered by our lack of data on non-Medicaid prescriptions as well as the secular trend in general prenatal vitamin use.

Given the improved receipt of prenatal care that occurred with the implementation of presumptive eligibility, why did we not observe corresponding reductions in the rates of moderately low birthweight, very low birthweight, and perinatal, neonatal, and infant mortality? According to an Institute of Medicine report, prenatal care improves pregnancy outcome by first reducing the occurrence of low-birthweight deliveries.[1] This reduction is, in turn, reflected in improved survival of neonates. Such results may become evident in studies with larger numbers of participants or longer times from intervention to assessment than were the case in the present study, or adjustments in the content of prenatal care may be required.

Several factors limit these findings. First, the analysis is restricted to a single state, Tennessee. The situation may be different in other states. Second, there may be confounding by secular trend; that is, there may be a tendency independent of the Medicaid expansions for women to obtain prenatal care earlier and more regularly over time. The pre- and postintervention nature of our study design limits our ability to detect such a trend. Third, our data sources (vital records and Medicaid enrollment files) may not contain sufficient information to detect important outcome changes. For example, these sources do not contain information on the content and quality of prenatal care. Fourth, subsequent Medicaid regulatory changes, namely enactment of expanded eligibility to pregnant women living at 150% of the federal poverty level on January 1, 1990, permitted only a short time from enactment of presumptive eligibility to our evaluation of its effects. It is possible that as clients and health care providers become more familiar with presumptive eligibility, it will have a much greater effect on perinatal outcome.

We believe that this analysis of presumptive eligibility is the first study of Medicaid expansion to demonstrate a consistent improvement in process, if not outcome. When barriers to prenatal care, including bureaucratic ones, are removed, low-income women will seek care earlier and more frequently. Other recent studies of expanded prenatal services for pregnant women have reported either very limited[17] or no success[18] in terms of improved receipt of care and reduction of adverse outcomes. These studies, along with the data presented here, suggest that universal health care coverage for pregnant women will improve the receipt of prenatal care. Whether prenatal care alone is sufficient to improve the perinatal outcome for low-income women remains to be demonstrated. □

Acknowledgments

This study was funded by Health Care Financing Administration Cooperative Agreement No. 18-C-90029/4-02.

References

1. Institute of Medicine. *Preventing Low Birthweight.* Washington, DC: National Academy Press; 1985.
2. Greenberg RS. The impact of prenatal care on different social groups. *Am J Obstet Gynecol.* 1983;145:797–801.
3. Murray JL, Bernfield M. The differential effect of prenatal care on the incidence of low birth weight among blacks and whites in a prepaid health care plan. *N Engl J Med.* 1988;319:1385–1391.
4. *Caring for Our Future: The Content of Prenatal Care—A Report of the Public Health Service Expert Panel on the Content of Prenatal Care.* Washington, DC: US Dept of Health and Human Services, Public Health Service; 1989.
5. Centers for Disease Control and Prevention. Recommendations for the use of folic acid to reduce the number of cases of spina bifida and other neural tube defects. *MMWR.* 1992;41(No. RR-14):1233–1238.
6. *Prenatal Care: Early Success in Enrolling Women Made Eligible by Medicaid Expansions.* Washington, DC: US General Accounting Office; 1991.
7. Piper JM, Ray WA, Griffin MR. Effects of Medicaid eligibility expansion on prenatal care and pregnancy outcome in Tennessee. *JAMA.* 1990;2219–2223.
8. Piper JM, Mitchel EF Jr, Ray WA. Expanded Medicaid coverage for pregnant women to 100 percent of the federal poverty level. *Am J Prev Med.* In press.
9. Breyel J. *State Coverage of Pregnant Women and Children: July 1991.* Washington, DC: National Governors Association; 1991.
10. Breyel J, Curley T. *State Coverage of Pregnant Women and Children: January 1992.* Washington, DC: National Governors Association; 1992.
11. Ray WA, Griffin MR. Use of Medicaid data for pharmacoepidemiology. *Am J Epidemiol.* 1989;129:837–849.
12. Piper JM, Ray WA, Griffin MR, et al. Methodological issues in evaluating expanded Medicaid coverage for pregnant women. *Am J Epidemiol.* 1990;132:561–571.
13. *Tennessee Statistical Abstracts.* Knoxville, Tenn: University of Tennessee, Center for Business and Economic Research; 1991.
14. Kessner DM, Singer J, Kalk CE, et al. Infant death: An analysis by maternal risk and health care. In: Institute of Medicine. *Contrasts in Health Status.* Washington, DC: National Academy Press; 1973.
15. Piper JM, Mitchel EF Jr, Snowden M, et al. Validation of 1989 Tennessee birth certificates using maternal and newborn hospital records. *Am J Epidemiol.* 1993;137:758–768.
16. Dixon WJ, Brown MB, Engleman L, et al. *BMDP Statistical Software Manual.* Vols. 1 and 2. Berkeley, Calif: University of California Press, 1990.
17. Braveman P, Bennett T, Lewis C, et al. Access to prenatal care following major Medicaid eligibility expansions. *JAMA.* 1993;269:1285–1289.
18. Haas JS, Udvarhelyi IS, Morris CN, et al. The effect of providing health coverage to poor uninsured pregnant women in Massachusetts. *JAMA.* 1993;269:87–91.

The Danger of Applying Uniform Clinical Policies across Populations: The Case of Breast Cancer in American Indians

ABSTRACT

Objectives. This study examined the implications of annual screening mammography for cost and mortality in American Indian populations with differing baseline breast cancer rates.

Methods. A decision tree compared annual screening mammography and screening clinical breast examination with referral for diagnostic mammography when appropriate The decision tree was constructed to examine the effect of different baseline cancer rates, stage at diagnosis, and stage-specific survival. Outcomes included 5-year relative survival, deaths prevented at 5 years, cost per death prevented, and total costs.

Results. The findings suggest that the total cost of breast cancer is 3.6 times higher with the screening mammography program but results in a 27.9% reduction in breast cancer deaths over the first 5 years of the program. Both costs and deaths prevented are sensitive to the incidence of breast cancer in the population and are less favorable in the range of incidence seen in American Indians.

Conclusions. The cost and impact of a given strategy for cancer screening vary among communities with different disease incidence, stage at diagnosis, and stage-specific survival, as seen in American Indian populations. (*Am J Public Health.* 1994;84:1631–1636)

Paul A. Nutting, MD, MSPH, B. Ned Calonge, MD, MPH, Donald C. Iverson, PhD, and Larry A. Green, MD

Introduction

Over the last 5 years, a great deal of attention has been given to developing national clinical policies as a strategy to reduce practice variation, improve patient outcomes, and reduce the costs of health care.[1-3] Community-based studies in chronic disease are highly valued and used to inform clinical policy decisions, and they often influence clinical policy for populations well beyond those in which the original studies were conducted. While some attention is given to the effect of sociocultural differences on health and care-seeking behaviors, less attention is paid to the generalizability of results across populations with differing baseline disease rates.

The difficulty in developing a rational policy for breast cancer screening for the different populations served by the Indian Health Service (IHS) illustrates a universal but poorly appreciated dilemma. The IHS provides health and medical care services to more than a million American Indians and Alaska Natives in the continental United States and Alaska and allocates a finite resource base across a broad range of strategies for improving the health status of these populations. Allocation decisions involve trading off the potential benefits of a particular strategy against others for the same or competing health problems. To complicate clinical policy decisions for the IHS, there is considerable variation in the occurrence of different health problems among different American Indian and Alaska Native populations, as well as variation in the probability of success for the potential strategies to deal with these problems. Moreover, there is often variation in the cost of a given strategy in its application to the vastly different socioeconomic and political environments of American Indian and Alaska Native communities.

Screening for breast cancer presents an important clinical policy question for the IHS. The incidence of breast cancer in American Indian and Alaska Native populations is believed to vary from 20% to nearly 70% of the rate in the US general population.[4-9] There is evidence that American Indian and Alaska Native women with breast cancer are diagnosed at a more advanced stage of disease and that stage-specific survival is worse than for non-Indian women.[10] This study used formal decision analysis to examine the cost and outcome implications of a given clinical policy for screening mammography when applied to communities with differing breast cancer rates, stage at diagnosis, and stage-specific survival experiences.

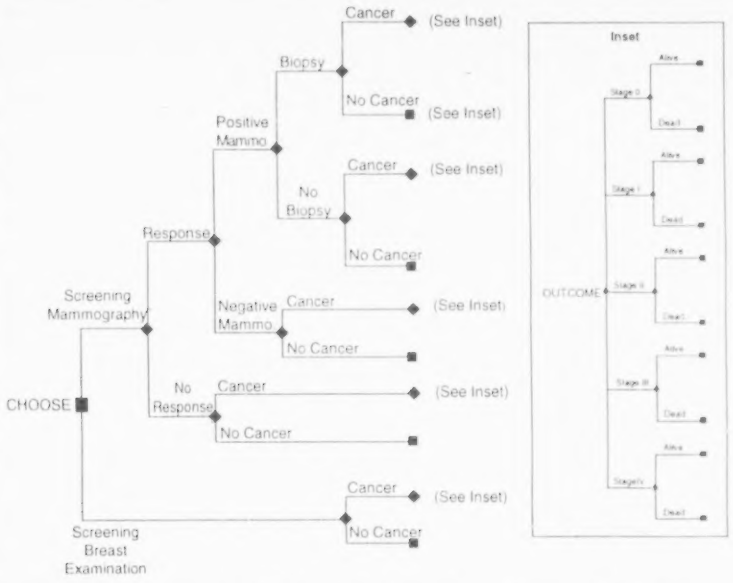
Methods

A decision tree was constructed to compare two strategies for breast cancer screening (Figure 1) and analyzed with SmlTree. The first strategy (screening mammography) consisted of annual

The authors are with the Department of Family Medicine, University of Colorado Health Sciences Center, Denver. Paul A. Nutting and Donald C. Iverson are also with the Ambulatory Sentinel Practice Network, Denver. At the time this article was written, Paul A. Nutting was with the Indian Health Service, Tucson, Ariz.

Requests for reprints should be sent to Paul A. Nutting, MD, MSPH, Department of Family Medicine, University of Colorado Health Sciences Center, 1180 Clermont St, Denver, CO 80220.

This paper was accepted January 26, 1994.



Note. The analysis compared two strategies for controlling breast cancer: (1) annual screening mammography for all women 50 years of age and older and (2) routine clinical breast examination. Data used for the clinical breast examination strategy reflected incidence of breast cancer, current levels of screening, and breast cancer outcomes in the population. Breast cancer outcomes are shown in the inset and include stage at diagnosis and stage-specific 5-year survival.

FIGURE 1—The decision tree: two strategies for controlling breast cancer.

screening mammography and clinical breast examination in women 50 years of age and older. Consistent with logistic considerations in the IHS, this strategy is based on the use of mobile mammography units with intensive community outreach to encourage use of the screening opportunity. The second strategy (clinical breast examination) reflects current screening practices in most IHS programs, which rely largely on periodic clinical breast examination followed by diagnostic mammography for suspicious masses. Current data for incidence, stage at diagnosis, and stage-specific survival were used to model the second strategy as described below.

The analysis consisted of three components. First, the base analysis compared the two screening strategies for the (1) distribution of stage at diagnosis, (2) specific 5-year relative survival, (3) deaths prevented at 5 years, (4) cost per death prevented, and (5) total costs of breast cancer. In order to put the results of the study into a community perspective, the results of the decision analysis were applied to the age and gender distribution of a specific American Indian community in the southwestern United States. The community used is typical in size and age–sex distribution of many reservation-based communities served by the IHS. It consists of 13 620 people, with 2 888 women 25 years of age and older. The model expressed the effect of the two screening strategies on the 775 women in this population 50 years of age and older.

A second analysis examined the sensitivity of the outcomes applied to the study community to critical assumptions of the analysis, including the cost of mammography, the cost of diagnostic workup of patients with positive screening tests, and patient compliance in obtaining mammograms.

A third analysis examined the effect of the underlying incidence of disease in the population on the outcomes of the decision analysis. This was accomplished by varying the rate of breast cancer in the analysis using incidence figures for American Indian women in Arizona and New Mexico, Alaska Native women, and White women in New Mexico. The results of the analysis for each group were applied to the age and gender distributions of the study community.

Estimates of Incidence, Stage at Diagnosis, and Survival

Estimates of the annual incidence, the stage at diagnosis, and 5-year survival were obtained from the population-based New Mexico Tumor Registry for the years 1973 through 1988. Annual incidence rates for breast cancer for each of four age groups (25 through 39 years, 40 through 49 years, 50 through 64 years, and 65 through 89 years) for Navajo and non-Navajo Indian women in Arizona and New Mexico, White women in New Mexico, and Alaskan Native women were computed and age adjusted to the 1970 US population.

The analysis used estimates of the distribution of stage at diagnosis of breast cancer with and without a screening mammography program developed by the National Cancer Institute Working Group on Cancer Screening and Detection.[11] These estimates were based on the 30% reduction in breast cancer mortality for women 50 to 70 years of age reported in the literature. Criteria established by the American Joint Committee on Cancer[12] were used to compute the observed stage at diagnosis for Navajo and non-Navajo Indian women in Arizona and New Mexico from New Mexico Tumor Registry data.

The Kaplan-Meier product-limit method was used to produce a survival function from the registry data. Cases that were not known to involve death were censored at the date of last documented follow-up. Stage-specific survival rates at 5 years were obtained from the survival function and computed separately for Navajo and for all other Arizona and New Mexico Indians (tables showing the age-specific estimates of incidence, distribution of stage at diagnosis, and stage-specific survival rates at 5 years for Navajo and non-Navajo American Indian women in Arizona and New Mexico used in the analysis are available on request from the authors).

Costs and Assumptions Used in the Analysis

The assumptions incorporated into the analysis that dealt with cost, participation rates, and sensitivity and specificity of the screening and diagnostic procedures are summarized in Table 1. The analysis assumed a 50% response and compliance with annual screening mammography over 5 years, consistent with the rate for all scheduled screenings in the Health Insurance Plan study[13] but lower than the

initial response reported from the Group Health Cooperative of Puget Sound study.[14] The analysis also assumed that those individuals responding to the invitation for screening were of the same risk level and thus had the same distribution of cancer by stage as those not responding. Finally, the analysis assumed that 95% of women would comply with the diagnostic workup that followed positive mammography.[11]

The sensitivity and specificity of screening mammography were assumed to be 75% and 94%, respectively, based on reports from the Canadian National Breast Screening Study.[15] The primary analysis also assumed that the sensitivity and specificity of surgical biopsy were both 100%. The probability of a positive screening test is a function of the prevalence of undetected breast cancers in the population presenting for screening and the sensitivity and specificity of the screening strategy used. The prevalence of undetected cancers was estimated from the annual incidence of breast cancer in the population, with the recognition that this approximation may underestimate prevalence in the initial years of the screening program when prevalent (but previously undiagnosed) cases are found.

The primary analysis assumed a cost of $30 per mammogram including interpretation, a rate that was under negotiation for another study within the IHS. Reported estimates range from $20 to $200,[16-19] and sensitivity analysis examined the impact of varying this estimate from $20 to $120. The cost of clinical breast examination has been estimated from $10 to $50[18] and as low as $3.25 when conducted during a regular examination rather than during a separate visit.[17] For the primary analysis, it was assumed that women receiving screening mammography would receive a clinical breast examination at the same time, at a cost of $3.25. The primary analysis also included a cost of outreach of $25 per patient participating in the mammography screening program. This was estimated to be the per-patient cost associated with the community health worker time required to publicize the program and the common practice of providing transportation for many patients to the health care facility. Estimates of biopsy costs vary widely, ranging from $500[20] to nearly $2000.[21] A study in the Alaska area of IHS estimated the cost of biopsy at $100 to $500.[22] For simplicity, the analysis assumed an average cost of $500 to include surgical biopsy and other diagnostic procedures such as additional mammographic views, needle aspiration, or ultrasound examinations that may be ordered. Sensitivity analysis examined the effect of varying this estimate from $500 to $1500. Estimates of the cost of care for breast cancer by stage of the disease and for terminal care were modified from those of Eddy,[18] which were based on Medicare data.

TABLE 1—Assumptions Used in the Primary Analysis

Screening mammography with breast self-examination	
Sensitivity	0.75
Specificity	0.94
Response for annual screening, %	
Mammography	50
Compliance with referral for breast examination	95
Compliance with referral for diagnostic mammography	95
Cost of screening and diagnosis, $	
Mammography	30.00
Biopsy	500.00
Cost of outreach	25.00
Cost of clinical breast exam (done at time of mammography)	3.25
Cost of care for cancer patients, $	
Stage 0	7 970
Stage 1	8 219
Stage 2	8 809
Stage 3	8 809
Stage 4	11 859
Terminal care	24 500

Source. Cost of care data are derived from Medicare data and reported by Eddy.[18] (Reproduced, with permission, from Eddy.[18])

TABLE 2—Baseline 5-Year Cancer Deaths and Estimated Cost of Care for Cancer Patients in the Absence of a Mammography Screening Program in the Study Community

	Age, y					
	25–34	40–49	50–64	65 and Over	Total Over 25	Total Over 50
No. of women in age group	1 627	704	666	109	2 888	775
Annual incidence per 100 000 women	7.69	27.10	48.60	70.40		
Total cancer deaths in group	0.24	0.39	0.67	0.55	1.85	1.22
Cost of care for cancer, $	2 392	3 654	6 187	5 290	17 523	11 477

Results

Comparison of Outcomes

The "screening breast examination" arm of the decision tree examined the baseline cancer deaths and costs of care for breast cancer patients in the absence of a program of screening mammography. Table 2 shows these data for a 5-year period applied to several age categories of the study community. The results show relatively few breast cancer deaths (1.85) over a 5-year time frame in women 25 years of age and older; 34% of these occur in the age group 25 through 49, which, despite recommendations for women 40 to 49 years old, is generally not considered a target population for regular screening mammography.

In comparison, the "screening mammography" arm of the tree modeled the effect of an intervention that offered annual screening mammography and clinical breast examinations for all women 50 years of age and older. Table 3 summarizes the costs and effects of the mammography strategy when applied to the age distribution of the study community. Offer-

TABLE 3—5-Year Cumulative Costs and Effects of Mammography Screening for Breast Cancer in Women 50 Years of Age and Older in the Study Community

Cancer deaths expected	1.22
Cancer deaths experienced	0.88
Cancer deaths prevented	0.34
Reduction in deaths, %	27.9
Cost of screening mammography program, $	32 002
Cost of care for cancer patients, $	9 617
Total cost, $	41 619
Screening program cost per cancer death prevented, $	94 124

Note. Screening involves annual screening mammography of all women 50 years of age and older. n = 1059.

TABLE 4—5-Year Cumulative Costs and Outcomes of Screening in a Single Study Population Based on Assumptions of Baseline Incidence Rates as Observed in Different Populations

	Non-Navajo American Indians in Southwest 40–49 Years Old	Non-Navajo American Indians in Southwest 50 Years of Age and Older	Alaska Natives 50 Years of Age and Older	Whites in New Mexico 50 Years of Age and Older
No screening program				
Age-adjusted incidence (per 100 000)	27.1	56.5	116.6	294.5
Deaths expected with no screening program	0.59	1.22	2.49	6.27
Expected cost of care for cancer patients, $	5 439	11 480	23 616	59 632
Screening program				
Deaths prevented	0.14	0.34	0.68	1.72
Expected cost of care for cancer patients, $	4 829	9 617	19 804	50 039
Cost of screening program, $	32 002	32 002	32 002	32 002
Total cost, $	36 831	41 619	51 806	82 041
Cost of screening program per death prevented, $	228 585	94 124	47 062	18 606

Note. Incidence rates were age adjusted on 5-year intervals to the study population.

ing annual screening mammography to women 50 years old and older in this community over 5 years increases the cost of breast cancer care by a factor of 3.6 (from $11 477 in the baseline for cost of care of cancer cases to $41 619 for total program cost with the screening program). The cost is associated with a 27.9% decrease in cancer deaths (from 1.22 to 0.88) over the same 5 years. There is also a small savings in cost of care with screening: a decrease from $11 477 to $9 617, clearly not enough to offset the higher cost of the screening program itself. In comparison with the baseline, the cost of the screening mammography program is $94 124 for each cancer death averted over 5 years of program operation.

Sensitivity Analysis

Sensitivity analysis was conducted to examine the effect of the cost estimates for mammography used in the primary analysis. The findings suggest that the decision result is not very sensitive to the cost of mammography: a fivefold increase in cost of mammography (from $20 to $100) increases the cost per death prevented by only 10.1%.

Sensitivity analysis was also conducted for the cost of surgical biopsy and other diagnostic procedures, examining the effect of varying the cost from $500 (as used in the base analysis) to $1500, which increased the overall program cost by less than 10%.

Sensitivity analysis also examined the effect of varying the assumption of 50% compliance with screening mammography from 10% to 90% on the total cost, cancer deaths prevented, and cost per cancer death prevented. The total program cost over 5 years in the study community increases steadily (from $6400 to $58 354) with the response rate, consistent with the observation that the majority of the cost is associated with the screening and diagnostic procedures, with reductions in the cost of care for late stage and terminal cases being small in comparison. Assuming that the individuals at highest risk are neither the early nor the late responders, the number of deaths prevented decreases steadily with decreasing response rates. For example, at a 90% response rate, approximately 50% of baseline cancer deaths could be prevented in the age group 50 years and older. At a response rate of 67%, approximately 30% of baseline deaths would be prevented, consistent with the observations of most large-scale studies of screening mammography.

The effect of varying the compliance rate with screening on the rate of early detection of breast cancer was also examined; stage 0 or 1 at diagnosis was considered early detection.[17] Across the full range of response rates, annual screening mammography produces early detection rates from 18.4% to 58%.

Effect of Varying Incidence Rates

The decision to make screening mammography services widely available through the IHS is confounded by the varying incidence rates of breast cancer among different American Indian and Alaska Native communities. The effect on program outcome of different incidence rates of breast cancer was examined. The age-specific incidence rates for American Indians in Arizona and New Mexico, Alaska Natives, and White women in New Mexico were age adjusted on 5-year intervals to the study population to compute an age-adjusted incidence for women in each group 50 years of age and older and for American Indian women in Arizona and New Mexico 40 to 49 years old. The different incidence rates were then used in the model and applied to the study population. Table 4 compares program cost and outcome for the different

incidence rates. These data suggest that the cost per death prevented by screening decreases with increasing breast cancer incidence, while the benefit in terms of deaths prevented increases rather sharply even though screening produces a constant reduction in deaths of 20% to 30% at all incidence levels. Figure 2 shows the cost per cancer death prevented at different incidence rates. Clearly, the cost per death prevented not only is higher at the comparatively lower incidence rates common among American Indian communities but rises very steeply for low incidence rates. At the incidence observed in 40- to 49-year-old Indian women in the Southwest, the cost per death prevented by screening is 12 times higher than the cost at an incidence rate common in White women 50 years old and older.

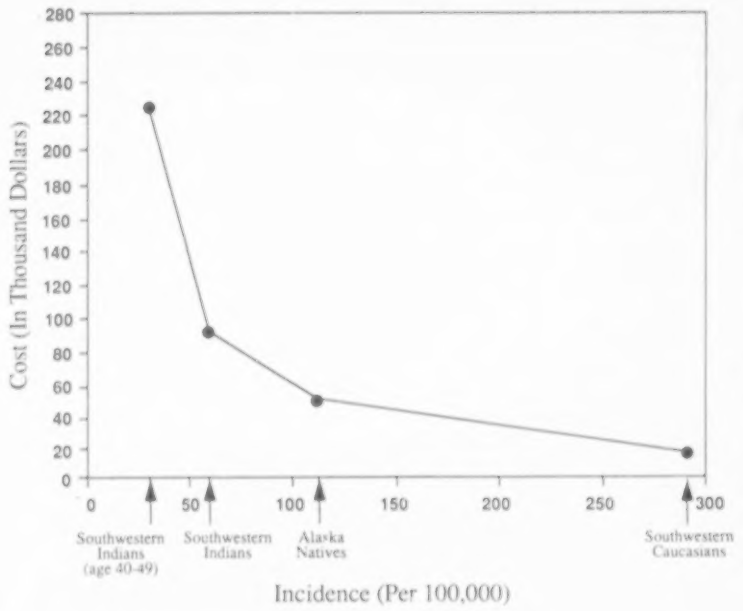


FIGURE 2—Cost per death prevented over 5 years as a function of baseline incidence of breast cancer in women 50 years of age and older.

Discussion

The primary analysis suggests that annual screening mammography could be expected to result in a 27.9% reduction in breast cancer deaths over 5 years. Applying the model output to a specific reservation-based Indian community with a low incidence of breast cancer and a youthful age distribution helps put program outcomes in perspective. For example, the baseline mortality reduction is small in absolute terms, since in the study population of 13 620 only 1.22 breast cancer deaths in women 50 years of age and older would be expected over 5 years in the absence of a screening program. Of these, 34% of deaths are expected in the age group under 50 years, a group that might not be targeted in a screening mammography program. Seen another way, offering screening mammography to this community would save a breast cancer life once in every 15 years at a program cost of $124 857 over the same period.

Since there are relatively few instances of breast cancer, the cost of care for cancer patients is relatively small in comparison with the cost of the screening program in many American Indian communities. The cost of cancer patient care projected for the study population over a 5-year time frame decreased by 16.2%, from $11 477 at baseline to $9617 with mammography. Although the cost of screening will not be substantially offset by reductions in the cost of care for patients with breast cancer in American Indian and Alaska Native communities, cost of cancer care exceeds cost of screening when dealing with the incidence levels of White women of the Southwest (Table 4).

Estimates of the cost-effectiveness of screening mammography have varied widely,[16,18,21,23-25] from a high of $126 000[18] to a low of $5600[25] per year of life saved when compared in 1991 dollars.[23] Our primary outcome measure, cost per cancer death prevented in the study community over 5 years of annual screening, is not easily translated to cost per year of life saved. When we apply our model to the cancer incidence observed in White women in the Southwest, however, we conclude that terminal care costs exceed the cost of the screening program and that the program costs $18 606 per death prevented. Both suggest more screening benefit than shown in previous studies. It should be noted that applying White incidence rates to our model does not constitute an analysis of program effect in a White community. All other assumptions specific to the American Indian and Alaska Native populations are retained, including a relatively late stage at diagnosis (in the absence of screening) and relatively poor stage-specific survival, both of which will magnify the benefit of screening mammography.

This study demonstrates the danger of applying a uniform clinical policy across populations with differing baseline disease rates, stage at diagnosis, and stage-specific survival. The analysis suggests that several program outcomes are sensitive to the changing incidence of breast cancer in women 50 years old and older, as might be seen across different ethnic populations or across international borders. Under base assumptions with results applied to the study community, the cost of the screening program per breast cancer death prevented rises sharply from $18 606 when using incidence rates observed in southwestern White women to $94 124 for rates from southwestern Indian women. Moreover, the cost per death prevented increases even further to $228 585 for the incidence rates observed among southwestern Indian women between the ages of 40 and 49 years.

The finding that cost and program outcome are related to baseline incidence in the population is consistent with fundamental principles of epidemiology. That the trends are as striking in the range of incidence rates observed in American Indian women, however, is of considerable policy importance. It underscores the point that a relatively greater impact can be achieved by addressing health problems that are relatively common rather than those that are rare. It also empha-

sizes the importance of targeting high-risk groups, particularly in populations with low baseline incidence, and thus increasing the prevalence of breast cancer in the group to be targeted with screening.

An adverse distribution of stage at diagnosis and poor stage-specific survival have been observed among many ethnic minorities.[10,26–28] As might be expected, the impact of screening mammography appears to increase in populations with worse distributions of stage at diagnosis, such as the Navajo population. When Navajo women were compared with other Indian women, the analysis suggested a further 26.4% reduction in 5-year mortality and a reduction in cost per death prevented from $94 124 to $78 054. These effects assume no change in stage-specific survival. The poor stage-specific survival, however, emphasizes the importance of considering other intervention strategies for breast cancer mortality reduction as well. For example, a simple analysis based on the stage-specific survival data used in the model suggests that important reductions in deaths at 5 years might be achieved (in the absence of screening mammography) by improving stage-specific survival to the level of other American Indian communities in Arizona and New Mexico (a 24.1% reduction) or to the level observed in Surveillance Epidemiology and End Result programs[29] for all races (a 42.4% reduction). The reduction possible by raising the survival to the level of other Indian communities is nearly the same as that achievable by screening mammography and deserves further study.

Conclusions

This study developed a decision model to examine the cost and outcomes associated with screening mammography in American Indian communities. Results suggest that cost per death prevented is very sensitive to baseline incidence and that the relationship is much less favorable in the range of incidence common in American Indian communities. The analysis also examined the relative impact of improving stage at diagnosis through improved screening compared with the impact of improving stage-specific survival through case management strategies. When applied to a population with an adverse distribution of both stage at diagnosis and stage-specific survival, the results suggest that important reductions in preventable deaths might be achieved through combinations of screening and case management. The results illustrate the potential effects of multiple intervention strategies and highlight the danger of applying a uniform clinical policy across populations with differing disease characteristics. □

Acknowledgment

We gratefully acknowledge the skilled assistance of Ms Lisa Preston in preparation of the graphic material for the manuscript.

References

1. Institute of Medicine. *Clinical Practice Guidelines: Directions for a New Program.* Washington, DC: National Academy Press; 1990.
2. Agency for Health Care Policy and Research. *Program Note: Clinical Guideline Development.* Rockville, Md: US Dept of Health and Human Services, Public Health Service; 1990.
3. Clinton JJ. From the Agency for Health Care Policy and Research: physician input invited on clinical guideline development. *JAMA.* 1991;265:1508.
4. Young TK, Frank JW. Cancer surveillance in a remote Indian population in northwestern Ontario. *Am J Public Health.* 1983;73: 515–520.
5. Sorem KA. Cancer incidence in the Zuni Indians of New Mexico. *Yale J Biol Med.* 1985;58:489–496.
6. Mahoney MC, Michalek AM, Cummings KM, Nasca PC, Emrich LJ. Cancer surveillance in a northeastern Native American population. *Cancer.* 1989;64:191–195.
7. Lanier AP, Bulkow LR, Ireland B. Cancer in Alaskan Indians, Eskimos, and Aleuts 1969–83: implications for etiology and control. *Public Health Rep.* 1989;104:658–664.
8. Norsted TL, White E. Cancer incidence among Native Americans of western Washington. *Int J Epidemiol.* 1989;18:22–27.
9. Nutting PA, Freeman WL, Risser DR, et al. Cancer incidence among American Indians and Alaska Natives, 1980 through 1987. *Am J Public Health.* 1993;83:1589–1598.
10. Samet JM, Key CR, Hunt WC, Goodwin JS. Survival of American Indian and Hispanic cancer patients in New Mexico and Arizona, 1969–82. *J Natl Cancer Inst.* 1987;79:457–463.
11. Levin DL, Gail MH, Kessler LG, Eddy DM. A model for projecting cancer incidence and mortality in the presence of prevention, screening, and treatment programs. *Natl Cancer Inst Monogr.* 1986;2:83.
12. Beahrs OH, Myers MH, eds. *Manual for Staging of Cancer.* 2nd ed. Philadelphia, Pa: Lippincott; 1983.
13. Chu KC, Smart CR, Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan Clinical Trial. *J Natl Cancer Inst.* 1988;80:1125–1132.
14. Taplin S, Anderman C, Grothaus L. Breast cancer risk and participation in mammographic screening. *Am J Public Health.* 1989;79:1494–1498.
15. Baines CJ, McFarlane DV, Miller AB. Sensitivity and specificity of first screen mammography in 15 NBSS centres. *J Can Assoc Radiol.* 1988;39:273–279.
16. Moskowitz M. Cost-benefit determinations in screening mammography. *Cancer.* 1987; 60(suppl):1680–1683.
17. Carter AP, Thompson RS, Bourdeau RV. A clinically effective breast cancer screening program can be cost-effective, too. *Prev Med.* 1987;16:19.
18. Eddy DM. Screening for breast cancer. *Ann Intern Med.* 1989;111:389–399.
19. Sickles EA, Wever WN, Galvin HB, Ominsky SH, Sollitto RA. Low-cost mammography screening: practical considerations with emphasis on mobile operation. *Cancer.* 1987;60:1688–1691.
20. Eddy DM. A computer-based model for designing cancer control strategies. *Natl Cancer Inst Monogr.* 1986;2:75.
21. Cyrlak D. Induced costs of low-cost screening mammography. *Radiology.* 1988;168: 661–663.
22. Mostow EN, Lanier A. Breast cancer screening in Alaska Native women. Anchorage, Alaska: Alaska Native Medical Center; 1989. (Available from Dr Lanier, Alaska Native Medical Center, Anchorage, AK).
23. Mushlin AI, Fintor L. Is screening for breast cancer cost-effective? *Cancer.* 1992; 69:1957–1962.
24. Okubo I, Glick H, Frumkin H, Eisenberg JM. Cost-effectiveness analysis of mass screening for breast cancer in Japan. *Cancer.* 1991;67:2021–2029.
25. van der Maas PJ, de Koning HJ, van Inevald BM, et al. The cost-effectiveness of breast cancer screening. *Int J Cancer.* 1989;43:1055–1060.
26. Dayal HH, Power RN, Chiu C. Race and socio-economic status in survival from breast cancer. *J Chronic Dis.* 1982;35:675–683.
27. Dayal HH, Polissar L, Dahlberg S. Race, socioeconomic status, and other prognostic factors for survival from prostate cancer. *J Natl Cancer Inst.* 1985;74:1001–1007.
28. Axtell LM, Myers MH. Contrasts in survival of Black and White cancer patients, 1960–73. *J Natl Cancer Inst.* 1978;60:1209–1215.
29. Horm JW, Asire AJ, Young JL, et al., eds. *SEER Program: Cancer Incidence and Mortality in the United States: 1973–81.* Bethesda, Md: National Cancer Institute; 1984. DHHS publication 85-1837.

Public Health Briefs

HIV Infection among Women in Prison: An Assessment of Risk Factors Using a Nonnominal Methodology

Catherine A. Hankins, MD, MSc, Sylvie Gendron, MSc, Margaret A. Handley, MPH, Christiane Richard, MD, Marie Thérèse Lai Tung, and Michael O'Shaughnessy, PhD

ABSTRACT

The relative contributions of needle use practices and sexual behaviors to human immunodeficiency virus (HIV) antibody seropositivity among 394 women incarcerated in Quebec were determined by risk factor assessment and serology with a nonnominal methodology. HIV positivity was found in 6.9% (95% confidence interval [CI] = 4.6, 9.9) of all participants and in 13% (95% CI = 8.6, 18.6) of women with a history of injection drug use. HIV seropositivity among women with a history of injection drug use was predicted by sexual or needle contact with a seropositive person, self-reported genital herpes, and having had a regular sexual partner who injected drugs, but it was not predicted by prostitution. Nonnominal testing is an ethical alternative to mandatory and anonymous unlinked testing among correctional populations. (*Am J Public Health.* 1994;84: 1637–1640)

Introduction

Individuals entering correctional institutions may have been at risk for acquiring human immunodeficiency virus (HIV) infection through injection drug use practices or unprotected sexual activity.[1,2] Among women inmates, prior injection drug use has been associated with HIV infection,[3] and those under age 25 appear more likely to be infected than similarly aged incarcerated men.[4]

Studies that detail previous risk behaviors can inform the design of interventions to reduce HIV transmission both in prisons and in the outside community. The objective of the first such study in Canada was to examine, with a nonnominal methodology, the relative contributions of preincarceration needle use and sexual practices to HIV antibody seropositivity among incarcerated women.

Methods

Study participation was inmate initiated and confidential. Freely given informed consent was obtained before private administration of a risk factor questionnaire and venipuncture. A numbered ticket chosen randomly from a box by the inmate allowed for the nonnominal linkage of questionnaire data to serological results. Pretest counseling followed to assist the inmate in determining whether she wished to learn her serostatus. On subsequent presentation of her number (while incarcerated, on release, or thereafter), a consenting inmate was informed of her results. No information was provided to the prison health service or administration. After posttest counseling, medical and psychosocial referrals were arranged on request.

Serological determinations for HIV antibodies were done with commercial enzyme-linked immunoassays, immunoblot, and/or radio immune precipitation assay following World Health Organization criteria.[5] Basic sociodemographic data were compiled on all admissions during 1988 and 1989 for prison stays of 1 week or longer.[6] Chi-square tests, Student's *t* tests, Fisher's exact tests, and forward stepwise logistic regression were used to examine the relationship between HIV seropositivity and risk factor data. Mantel–Haenszel estimators for odds ratios and Miettinen's test-based 95% confidence intervals (CIs) are reported.

Results

A total of 394 women participated, with Montréal residents ($P < .001$) and women of Black and other (nonaboriginal) origin ($P < .001$) overrepresented in the study (Table 1). Participants also had lower educational levels than those of the general prison population ($P < .001$).

Overall, 6.9% of participants, 13.0% of women with a history of injection drug use, and 12.9% of those for whom

Catherine A. Hankins, Sylvie Gendron, Margaret A. Handley, and Marie Thérèse Lai Tung are with the Centre for AIDS Studies, Montréal-Centre Regional Public Health Team, Montréal, Québec, Canada. Christiane Richard is with the CLSC Ahuntsic, Montréal. Michael O'Shaughnessy is with the Bureau of Laboratory and Research Services, Department of National Health Canada, Ottawa.

Requests for reprints should be sent to Catherine A. Hankins, MD, MSc, Montréal-Centre Regional Public Health Team, 1P8, 1616 boul René-Lévesque, Montréal, Québec, Canada H3H 1P8.

This paper was accepted January 26, 1994.

TABLE 1—Sociodemographic Data for Incarcerated Women

	Study Participants (n = 394)	Total Incarcerated Population (n = 1638)	P
Age, y			
Mean (range)	28.5 (18–64)	29.5 (18–66)	.01[a]
Median	27	28	
Montréal resident, %	62.2	51.1	<.001[b]
Ethnic/racial origin, %			
French or English	85.0	90.4	<.001[b]
American Indian, Inuit	4.8	3.4	Not significant
Black or other	9.9	3.5	<.001[b]
Missing	0.4	2.7	
Education, %			
Primary school only	33.5	13.4	<.001[b]
Some high school or greater	63.2	83.6	
Missing	3.3	3.0	

[a]Two-tailed *t*-test, Satterthwaite's approximation.
[b]Chi-square test.

TABLE 2—HIV-1 Serostatus among the Incarcerated Women of the Study

Characteristics	Seropositive No.	Seropositive %	95% CI
IDU	192	13.0[a]	8.6, 18.6
IDU nonprostitute	126	11.9[a]	6.8, 18.9
IDU prostitute	66	15.2[a]	7.5, 26.1
Non–IDU	202	1.0	0.1, 3.5
Non-IDU nonprostitute	183	0.5	0.0, 3.0
Non-IDU prostitute	19	5.3	0.1, 26.0
Prostitute overall	85	12.9[a]	6.6, 21.9
Total study population	394	6.9	4.6, 9.9

Note. CI = confidence interval; IDU = injection drug user.
[a]No statistically significant differences were found among these categories.

TABLE 3—Demographics and History of Study Sample, by Serostatus

	Seropositive (n = 27)	Seronegative (n = 367)	Odds Ratio	95% CI	P
Mean age, y	26.9	28.6		[a]	NS
High school attendance, %	66.7	60.1	1.1	0.4, 2.6	NS
Past history of any STD, %	74.1	60.5	1.9	0.8, 4.8	NS
Montréal resident, %	81.5	60.8	2.8	1.1, 7.4	.03
History of gonorrhea, %	48.2	28.1	2.4	1.1, 5.1	.03
History of genital herpes, %	18.5	3.5	6.2	2.3, 16.7	<.01[b]
History of any IDU since 1979, %	92.6	45.5	14.9	4.9, 46.1	<.001

Note. CI = confidence interval; STD = sexually transmitted disease; IDU = injection drug use; NS = not significant.
[a]Two-tailed *t* test; *t* = 1.6.
[b]Fisher's exact test.

prostitution had been the principal source of revenue before incarceration were seropositive for HIV antibodies (Table 2). Tables 3 and 4 present selected results of bivariate analyses. In light of the preponderance of injection drug use among seropositive women (25/27), further analysis was restricted to the 192 women who had injected drugs (Table 5).

In forward stepwise logistic regression, the variables "subject awareness of previous sexual or needle contact with a person with HIV infection" (adjusted odds ratio = 6.5, 95% CI = 2.6, 16.5; $P < .001$), "self-reported genital herpes" (adjusted odds ratio = 5.5, 95% CI = 1.5, 20.1; $P = .01$), and "having had a regular sexual partner since 1979 who ever injected drugs" (adjusted odds ratio = 2.7, 95% CI = 1.0, 7.6; $P < .05$) were found to be significantly associated with the probability of HIV-1 infection. Although the confidence interval for the last variable includes the value of 1, reducing the statistical certainty that its addition improves the existing model, a log likelihood ratio test comparing this last model with one including only the first two variables was significant. After including these three variables in the model, the factors "prostitution in the year pre-incarceration" and "number of years injecting" were not significantly associated with HIV infection.

Discussion

Few studies have examined the risk behaviors of incarcerated women.[3,7–9] This study indicates that injection drug use is the predominant risk factor for HIV infection among women inmates and that needle use practices contribute substantially to the risk of HIV infection among injection drug users. The significant contribution of self-reported *Herpes simplex* infection in logistic modeling supports the hypothesis that mucosal disruption increases susceptibility to viral acquisition[10] and emphasizes the importance of sexual risk factors. The selective contribution of the latter was difficult to differentiate because both needle sharing and unprotected sexual activity are likely to have occurred with at-risk partners.

Self-selection bias[9,11,12] may have led to an underestimation or overestimation of true HIV prevalence in this correctional institution. However, because the prevalence of reported risk behaviors was high and the sample size was large, participation bias is unlikely to have greatly affected HIV-associated risk be-

havior findings. Differential recall may have occurred for the time period under study, depending on the proximity of participation dates to prison entry.

Freely given consent with no penalties for participation or refusal is considered the most ethical approach to the study of HIV infection in correctional settings because it permits inmates to weigh the risks and benefits of participation.[13–15] The further step of isolating consent for research participation from pretest counseling procedures in this study allowed participants to freely decide whether or not to return for their serological results. The anonymity provided by voluntary nonnominal participation gave seropositive study subjects additional control of subsequent decisions to initiate medical and psychosocial follow-up. The nonnominal methodology employed in this study has research utility, permitting linkage of serological information to risk factor and demographic data, whether the latter is collected at prison entry, from inmate medical charts before testing, or through standardized interviews, as was done here.

Some studies comparing seroprevalence among prison entrants tested blindly with that among prison entrants volunteering under nonnominal or anonymous conditions have found no differences.[11,16] The introduction of voluntary testing can thus lead to increases in HIV detection, resulting in earlier initiation of prophylaxis and treatment.[17] Because some elements of disincentive are likely present in nominal testing even when confidentiality is ensured,[12] anonymous or nonnominal testing options can increase the effectiveness of voluntary testing in identifying inmates with HIV infection. In contrast to anonymous testing, which by definition must be provided by an outside team unaware of the identity of participants, nonnominal testing can be offered by correctional health services.

The correctional setting can be an important site for monitoring community drug use behavior, given the excellent availability of study volunteers with fewer time constraints. More injection drug users can be found in prisons than in drug treatment programs, hospitals, and social services.[18] In addition, reported preimprisonment behaviors related to drug use can be anticipated to be more heterogeneous among injection drug users studied in prison than among those studied in treatment facilities or at needle exchange programs.[19] Finally, the logistical ease of providing nonnominal voluntary testing combines with inmate compliance to challenge the belief that such testing cannot be provided in a prison setting. HIV-1 testing programs should be expanded to include voluntary, confidential, and anonymous testing, or failing this, nonnominal testing with incorporation of flexible follow-up options for the provision of test results. □

TABLE 4—Condom Use and Sexual Practices of Study Sample, by Serostatus

	Seropositive (n = 27)	Seronegative (n = 367)	Odds Ratio	95% CI	P
No. sexual partners in year before incarceration, median (range)	624 (0–10 000)	3 (0–10 000)	...[a]		.03
Any condom use since 1979, %	81.5	54.2	3.7	1.5, 9.5	< .01
Consistent (always) condom use with all partners for vaginal sex in past year,[b] %	11.5	6.4	1.9	0.5, 6.7	NS
Condom use (sometimes or always) with all partners for vaginal sex in past year,[b] %	61.5	39.5	2.5	1.1, 5.5	.03
Prostitution, %	40.7	20.2	2.7	1.2, 6.0	.01
IDU sexual partner,[c] %	88.9	44.7	9.9	3.6, 27.3	< .001
IDU regular partner,[d] %	77.8	31.3	7.7	3.4, 17.2	< .001
Seropositive sexual partner,[e] %	11.1	0.8	15.1	4.3, 53.7	< .005[f]

Note. CI = confidence interval; IDU = injection drug user; NS = not significant.
[a]Median two-sample test (normal approximation); $Z = 2.2$.
[b]Twenty-six of 27 seropositive women and 342 of 367 seronegative women had vaginal sex in the past year.
[c]Sexual partner since 1979 who has ever injected drugs.
[d]Regular partner in the 12 months before incarceration who ever injected drugs.
[e]Seropositive partner at any point since 1979.
[f]Fisher's exact test.

TABLE 5—Practices among Injection Drug Users of Study Sample, by Serostatus

	Seropositive (n = 25)	Seronegative (n = 167)	Odds Ratio	95% CI	P
No. years injecting					
Mean (range)	5.9 (1–19)	6.2 (< 1–21)	...[a]		NS
SD	4.5	5.1			
Median	5	5			
Prostitution, %	40.0	33.5	1.3	0.6, 3.1	NS
Needle borrowing/lending with a seropositive person, %	44.0	9.7[b]	7.3	3.1, 17.1	< .001
Needle borrowing/lending with a stranger, %	76.2[c]	44.3[d]	4.0	1.5, 10.9	.006

Note. CI = confidence interval; NS = not significant.
[a]Median two-sample test (normal approximation); $Z = .05$.
[b]For 165 seronegative respondents.
[c]For 21 seropositive respondents.
[d]For 140 seronegative respondents.

Acknowledgment

This study was supported by grant 6605-2951 from the National Health Research and Development Programme, Department of National Health Canada.

References

1. Statement from the Consultation on Prevention and Control of Aids in Prisons 1987. Geneva, Switzerland: World Health Organization; November 16–18, 1987.
2. Commissioner's Directive No. 821. Ottawa, Ontario: Canadian Correctional Services; January 1, 1988.
3. Smith PF, Mikl J, Truman BI, et al. HIV infection among women entering the New York State correctional system. *Am J Public Health.* 1991;81(suppl):35–40.
4. Vlahov D, Brewer TF, Castro KG, et al. Prevalence of antibody to HIV-1 among entrants to US correctional facilities. *JAMA.* 1991;265:1129–1132.
5. World Health Organization. Acquired immunodeficiency syndrome: proposed WHO

criteria for interpreting results from Western blot assays for HIV-1, HIV-2, and HTLV-I/HTLV-II. *Wkly Epidemiol Rec.* 1990;65:281–288.
6. Ministère de la Sécurité publique. Liste des admissions pour l'établissement de détention—1 janvier 1988 au 31 décembre 1989.
7. De Paula MDN, Lorenço R, Queiroz W, et al. Prevalence of HIV seropositivity and potential risk of vertical transmission at the Sao Paulo female state prison. Presented at the VIIth International Conference on AIDS, Florence, Italy, June 16–21, 1991. Abstract M.C.3002.
8. Granados A, Miranda MJ, Martin L. HIV seropositivity in Spanish prisons. Presented at the VIth International Conference on AIDS, San Francisco, Calif, June 20–24, 1990. Abstract Th.D.116.
9. Patel KK, Hutchinson C, Sienko DG. Sentinel surveillance of HIV infection among new inmates and implications for policies of correctional facilities. *Public Health Rep.* 1990;105:510–514.
10. Wasserheit JN. Epidemiological synergy: interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases. *Sex Transm Dis.* 1992;19:61–77.
11. Hoxie NJ, Vergeront JM, Frisby HR, Pfister JR, Golubjatnikov R, Davis JP. HIV seroprevalence and the acceptance of voluntary HIV testing among newly incarcerated male prison inmates in Wisconsin. *Am J Public Health.* 1990;80:1129–1131.
12. Andrus JK, Fleming DW, Knox C, et al. HIV testing in prisoners: is mandatory testing mandatory? *Am J Public Health.* 1989;79:840–842.
13. World Health Organization. Report on HIV/AIDS in Prison 1990. WHO Global Programme on AIDS and Crime Prevention and Criminal Justice Branch of the United Nations Office; Vienna, Austria; September 1990.
14. Hammett TM, Dubler NN. Clinical and epidemiologic research on HIV infection and AIDS research among correctional inmates—regulations, ethics and procedures. *Evaluation Review.* October 1990;14:482–501.
15. Dubler NN, Sidel VW. On research on HIV infection and AIDS in correctional institutions. *Milbank Q.* 1989;67:171–207.
16. Bird AG, Gore SM, Jolliffe DW, Burns SM. Anonymous HIV surveillance in Saughton Prison, Edinburgh. *AIDS.* 1992;6:725–733.
17. Kendig N, Behrendt C, Swetz A, Stough T, Coplin M, Vlahov D. Profile of seropositive inmates diagnosed in Maryland prisons. Presented at the VIIIth International Conference on AIDS, Amsterdam, the Netherlands, July 19–24, 1992. Abstract Po.D.5062.
18. Brewer TF, Derrickson J. AIDS in prison: a review of epidemiology and preventive policy. *AIDS.* 1992;6:623–628.
19. Vlahov D, Polk BF. Intravenous drug use and human immunodeficiency virus (HIV) infection in prison. *AIDS Public Policy J.* 1988;3:42–46.

Risk Factors for Acute Non-A, Non-B Hepatitis and Their Relationship to Antibodies for Hepatitis C Virus: A Case–Control Study

Alfonso Mele, MD, Luciano Sagliocca, MD, Giuseppe Manzillo, MD, Francesco Converti, MD, Pietro Amoroso, MD, Maria Antonietta Stazi, MSc, Luigina Ferrigno, Mariella Rapicetta, MD, Elisabetta Franco, MD, Brunella Adamo, MD, Filippo Palumbo, MD, Costanza Sbreglia, MD, Augusto Panà, MD, Paolo Pasquini, MD, and the SEIEVA Collaborating Group

ABSTRACT

A case–control study was carried out comparing 333 case subjects with non-A, non-B hepatitis and 1095 hospital control subjects. Of 333 case subjects, 197 (59%) were positive for hepatitis C antibody (anti-HCV). Excluding blood transfusion and intravenous drug use, surgical intervention and dental therapy were strongly associated with anti-HCV–positive cases: in particular, obstetric and gynecology surgical intervention was found to be strongly associated with HCV positivity (odds ratio [OR] = 32; 95% confidence interval [CI] = 7, 147). Raw shellfish consumption was a risk factor for anti-HCV–negative cases (OR = 2.2; 95% CI = 1.0, 5.1), thus suggesting an enterically transmitted virus in sporadic non-A, non-B hepatitis in Italy. (*Am J Public Health.* 1994;84:1640–1643)

Introduction

Non-A, non-B hepatitis includes several agents that are both blood and enterically transmitted.[1–3] A large number of patients with blood-transmitted non-A, non-B hepatitis may develop chronic liver disease. Blood transfusions and intravenous drug use explain fewer than 50% of the hepatitis cases; therefore, identification of additional risk factors for the disease is necessary.[4,5]

Risk factors associated with non-A, non-B hepatitis were evaluated in a case–control study based in three Naples hospitals (Italy) from November 1987 through May 1991.

Study Population and Methods

All case subjects were recruited in an infectious-diseases hospital, had serum

Alfonso Mele, Maria Antonietta Stazi, Luigina Ferrigno, and Paolo Pasquini are with the Istituto Superiore di Sanità, Laboratorio di Epidemiologia e Biostatistica, Roma. Luciano Sagliocca, Brunella Adamo, and Filippo Palumbo are with the Regione Campania, Osservatorio Epidemiologico, Napoli. Giuseppe Manzillo, Pietro Amoroso, and Costanza Sbreglia are with the Ospedale Cotugno, Napoli. Francesco Converti is with the Regione Campania, Servizio Epidemiologico USL 40, Napoli. Mariella Rapicetta is with the Istituto Superiore di Sanità, Laboratorio di Virologia, Roma. Elisabetta Franco is with the Cattedra di Igiene, Università della Basilicata, Potenza. Augusto Panà is with the Cattedra di Igiene, Università "Tor Vergata," Roma. For a list of the SEIEVA Collaborating Group, see the Acknowledgments.

Requests for reprints should be sent to Alfonso Mele, MD, Istituto Superiore di Sanità, Laboratorio di Epidemiologia e Biostatistica, Reparto di Epidemiologia Clinica, Viale Regina Elena 299, 00161 Roma Italy.

This paper was accepted December 23, 1993.

aminotransferase levels at least three times greater than the upper limit of normal values, and were negative for immunoglobulin M (IgM) antibodies to hepatitis A virus and hepatitis B core antigen (radioimmunoassay method, Abbott Laboratories). No case subject had a history or clinical sign of hepatic injury due to medication, alcohol, or other hepatobiliar diseases, and case subjects had an onset of disease not longer than 2 weeks.

Non-A, non-B case subjects with IgM antibodies against cytomegalovirus (Abbott CMV-M enzyme immunoassay) or with IgM antibodies against Epstein–Barr virus (Ortho monolert) were excluded from the study.

Serum samples from all case subjects were stored at −80°C.

Assays for hepatitis C antibodies (anti-HCV) were done before 1990 with the first-generation enzyme-linked immunosorbent assay (ELISA) (Ortho Diagnostics). Thereafter, the second-generation test was applied to all new cases and to serum samples that were negative for anti-HCV by the first-generation ELISA. Serum samples that showed an optical density between 0.7 and 1.7 were retested by supplemental second-generation recombinant immunoblot assay (RIBA) and were considered positive when two or more bands were present.

All anti-HCV–negative case subjects recruited from 1990 were followed-up for at least 6 months (range, 6–15 months).

The following control subjects were chosen: (1) patients without bloodborne and enterically transmitted disease admitted to the same hospital as the case subjects; (2) patients from the neurosurgery and ear, nose, and throat wards of a pediatric hospital; and (3) patients from the emergency, orthopedics, and surgery wards of a general hospital. Both the pediatric and general hospitals were located in the same area as the infectious-diseases hospital.

Case and control subjects lived in the same area and were interviewed on admission to the hospitals. Data for children were obtained from the mothers.

Exposure within 6 months and 6 weeks of the onset of disease was investigated for parenteral risk factors (including sexual exposure) and raw shellfish consumption, respectively, by means of a standardized questionnaire.

Age groups used for analysis were 0–9, 10–14, 15–19, 20–34, 35–54, and ≥ 55 years.

TABLE 1—General Characteristics of Control and Case Subjects, Naples, Italy, 1987 through 1991

	Subjects, %		
		Case	
Age, y	Control (n = 1095)	Positive for Hepatitis C Virus Antibody (n = 197)	Negative for Hepatitis C Virus Antibody (n = 136)
0–9	18.9	1.0	13.2
10–14	8.8	0.5	12.5
15–19	13.6	14.2	19.1
20–34	20.2	64.5	39.0
35–54	20.8	11.7	14.0
55 and over	17.7	8.1	2.2

TABLE 2—Diagnoses of Control Subjects (n = 1095)

Diagnosis	Percentage
Infectious diseases (Measles, mumps, brucellosis, rickettzial diseases, respiratory tract infections, encephalitis, herpes zoster, varicella, mononucleosis, toxoplasmosis)	13.3
Diseases following trauma and emergency surgery (Fractures, dislocations, abdominal trauma, acute appendicitis, cranial trauma)	22.9
Diseases undergoing surgical treatment (Orthopedic surgery, anorectal diseases, thyroid diseases, breast benign diseases, inguinal hernia)	32.1
Ear, nose, and throat diseases	9.7
Others (Inflammatory gastrointestinal diseases, depression, or arterial peripheral diseases, respiratory diseases, renal colic, irritable bowel syndrome)	21.9

Note. Two hundred twenty-one patients were recruited in the infectious-diseases hospital.

The polychotomous logistic technique was used to estimate the independent effect of the studied exposure variables on the risk of being either anti-HCV positive or anti-HCV negative.[6,7]

We first considered intravenous drug exposure, blood transfusion before and after 1990, age, sex, and education and thereafter the effects of all other exposure variables.

First-order interaction terms were always tested.

Results

Three hundred and forty-two case subjects and 1095 control subjects were recruited. Of the 342 case subjects, 251 (73%) were jaundiced; serum samples were taken from all case subjects within 21 days of the onset of disease. Four case subjects positive for IgM antibodies against cytomegalovirus and five positive for IgM antibodies against Epstein–Barr virus were excluded.

Of 333 case subjects, 187 (56%) were anti-HCV positive by the first- or second-generation test at first blood drawn (70 of 148 negative samples at first-generation ELISA were positive at second-generation ELISA and two of eight serum samples with an optical density between 0.7 and 1.7 were confirmed by second-generation RIBA).

Among the 68 anti-HCV–negative subjects followed-up, 10 (15%) became anti-HCV positive (second-generation test, second-generation RIBA confirmed): 5 subjects after 45 days from the onset of disease, 3 subjects after 2 months, and 2 subjects after 4 months. Overall, 164 of 248 (66%) males and 33 of 85 (39%) females were anti-HCV positive.

TABLE 3—Case and Control Subjects Exposed to Intravenous Drug Use and Blood Transfusion, Naples, Italy, 1987 through 1991

	Control Subjects		Case Subjects					
			Anti-HCV Positive			Anti-HCV Negative		
Risk Factor	%	Total No.[a]	%	Total No.[a]	OR (95% CI)	%	Total No.[a]	OR (95% CI)
Intravenous drug use	0.7	1086	51.3	195	71.0 (32.0, 160.0)	13.3	135	13.0 (5.3, 31.0)
Blood transfusion (1987–1989)	0.8	759	14.8	128	49.0 (17.0, 140.0)	4.6	87	7.0 (1.9, 27.0)
Blood transfusion (1990–1991)	0.3	333	1.4	69	1.9 (0.01, 270.0)	0.0	48	Not calculable

Note. Adjusted ORs (95% CIs) were estimated by polychotomous logistic regression according to positivity for hepatitis C virus antibody (anti-HCV). ORs were adjusted for age, sex, education level, and the other listed variables.
[a]Discrepancies in the numbers are due to missing values ("I don't know" answer).

TABLE 4—Case and Control Subjects Exposed to Considered Risk Factors, after Subjects with Blood Transfusions and Intravenous Drug Users are Excluded, Naples, Italy, 1987 through 1991

	Control Subjects		Case Subjects					
			Anti-HCV Positive			Anti-HCV Negative		
Risk Factor	%	Total No.[a]	%	Total No.[a]	OR (95% CI)	%	Total No.[a]	OR (95% CI)
Surgical intervention	2.4	1077	16.5	79	12.0 (4.0, 35.0)	12.4	113	4.9 (2.0, 12.0)
Hospitalization	10.3	1079	22.8	79	0.8 (0.3, 2.0)	17.5	114	1.1 (0.6, 2.3)
Dental therapy	11.0	1079	22.4	76	1.9 (1.0, 3.5)	13.2	114	1.2 (0.6, 2.1)
Other percutaneous exposures[b]	2.1	1072	5.1	78	1.7 (0.5, 6.0)	5.4	111	1.8 (0.7, 4.8)
More than one sexual partner[c]	6.6	775	16.9	77	1.4 (0.7, 3.0)	11.5	78	1.2 (0.5, 2.7)
Raw shellfish	5.7	715	16.7	54	1.4 (0.6, 3.6)	13.6	66	2.2 (1.0, 5.1)

Note. Adjusted ORs (95% CIs) were estimated by polychotomous logistic regression according to positivity for hepatitis C virus antibody (anti-HCV). ORs were adjusted for age, sex, education level, and the other listed variables.
[a]Discrepancies in the numbers are due to missing values ("I don't know" answer).
[b]Ear piercing, tattooing, acupuncture, electrolysis, or attendance at a chiropodist or manicurist.
[c]Age greater than 14 years.

In Tables 1 and 2 general characteristics of case and control subjects are presented: 26% of the anti-HCV–negative case subjects vs 1.5% of the anti-HCV–positive ones were younger than 15 years.

Among the case subjects, 118 were intravenous drug users and 24 received blood transfusions. The odds ratios were much higher for both these risk factors among anti-HCV–positive case subjects than among anti-HCV–negative ones (Table 3).

Of 215 case subjects, 23 (11%) reported receiving blood transfusions from 1987 through 1989; 1 of 117 case subjects (0.9%) received a blood transfusion after 1989.

Two anti-HCV–positive case subjects who were intravenous drug users received blood transfusions, 3 had surgery, and 43 had multiple sexual partners. Of the 20 anti-HCV–positive case subjects who received blood transfusions, 15 also had surgery.

After excluding blood transfusion and drug use, surgical intervention was significantly associated in both case series, and the risk was higher in anti-HCV–positive case subjects than in the anti-HCV–negative ones (Table 4). The strongest association was found in female case subjects age 15–44 years for obstetric and gynecology interventions: odds ratio (OR) = 32.0 (95% confidence interval [CI] = 7.5, 147.0) for anti-HCV–positive cases; OR = 16.5 (95% CI = 4.0, 73.0) for anti-HCV–negative cases. When control subjects undergoing surgical treatment were excluded, the ORs were 18 (95% CI = 5.1, 66.0) among anti-HCV–positive cases and 7.5 (95% CI = 2.6, 22.0) among anti-HCV–negative cases.

Dental therapy was associated with anti-HCV–positive cases, and a nonsignificant association was found between other percutaneous exposures and both cases series. A small but not significant excess of risk (OR = 1.4; 95% CI = 0.7, 3.0) was found among anti-HCV–positive case subjects exposed to multiple heterosexual partners (Table 4).

Thirty-three of 197 (17%) anti-HCV–positive case subjects and 66 of 136 (49%) anti-HCV–negative case subjects did not report any considered parenteral exposure.

A significant association existed between raw shellfish consumption and anti-HCV–negative cases. Raw shellfish consumption was the only risk factor for 7 of 66 (11%) anti-HCV–negative case subjects vs 3 of 54 (6%) anti-HCV–positive case subjects (Table 4).

No effect modification was found between the considered risk factors.

Discussion

More than 50% of the case subjects with non-A, non-B hepatitis were anti-HCV positive, but the diagnosis of acute hepatitis C is reliable only for the 10 case

subjects of 68 (15%) who seroconverted during the follow-up.

Blood transfusion and intravenous drug use account for 61% of the anti-HCV–positive cases and 16% of the anti-HCV–negative cases. Surgical intervention, dental therapy, and other percutaneous exposures are additional parenteral risk factors found to be associated with non-A, non-B hepatitis. Interestingly, the strongest association was shown with obstetric and gynecological interventions. Given that a large proportion of the general population is exposed to these risk factors, the present results and those reported for hepatitis B[8] underscore the importance of implementing efficient procedures for sterilization of instruments and the use of disposable materials, especially when a high turnover of patients combined with emergency surgical intervention (such as in gynecology) represents an increased risk of infection.

Few case subjects with non-A, non-B hepatitis and a history of blood transfusion were enrolled after 1989, when the anti-HCV test was introduced in the screening of blood banks. That fact suggests that a high number of non-A, non-B cases of hepatitis in patients reporting a history of blood transfusion were due to hepatitis C virus.

Heterosexual transmission of hepatitis B, non-A, non-B, and C viruses has been outlined in previous studies.[9–11] Our data show a slight (OR = 1.4; 95% CI = 0.7, 3.0) but not significant excess of risk for anti-HCV–positive case subjects exposed to multiple sexual partners.

The estimated odds ratios for blood transfusions, intravenous drug use, surgical intervention, and dental therapy were higher in the anti-HCV–positive cases. Although the anti-HCV test applied in this study does not discriminate acute from past infection, anti-HCV positivity elicits a group of non-A, non-B case subjects older than the anti-HCV–negative case subjects, and with greater exposure to parenteral risk factors. At present no test for additional non-A, non-B parenteral viruses is available, but it is likely that the anti-HCV–positive group includes acute non-A, non-B cases due to other blood-transmitted and orofecally transmitted viruses. At the same time, the anti-HCV–negative group might include hepatitis C cases.

The association of raw shellfish consumption with anti-HCV–negative cases suggests the possible presence of enterically transmitted non-A, non-B hepatitis in Italy.

Case subjects did not include all the patients with non-A, non-B hepatitis notified in the area, and hospital control subjects were selected in an attempt to avoid the selection process culminating with cases being diagnosed in the study hospital. The pediatric and general hospitals were located in the same area as the infectious-diseases hospital, where the control subjects would have been admitted had they been hospitalized with acute hepatitis. Furthermore, the broad diagnostic representation of control subjects probably prevented control selection bias.

The inclusion of patients undergoing surgical intervention as control subjects might have led to an underestimation of odds ratios: the associations between cases and surgical intervention were stronger after omitting these control subjects.

Control subjects with infectious diseases such as mononucleosis (n = 5) can share some risk factors with case subjects. When these control subjects were excluded, the estimate of odds ratios remained unchanged. □

Acknowledgments

This study was partially supported by Regione Campania, Assessorato alla Sanità, and by the National Research Council, Grant 91.00429.04.

We acknowledge Dr O. N. Gill of the Communicable Disease Surveillance Centre, P.H.L. London, for his contribution in setting up a national surveillance system of acute viral hepatitis (SEIEVA), of which this study is a part.

The SEIEVA Collaborating Group consists of select staff of the following institutions: the Regione Campania, Osservatorio Epidemiologico, Napoli; the Ospedale Cotugno, Napoli; the Regione Campania, Servizio Epidemiologico USL 40, Napoli; the Istituto Superiore di Sanità, Laboratorio di Virologia, Roma; and the Cattedra di Igiene, Università "Tor Vergata," Roma.

References

1. Tabor E. The three viruses of nonA, nonB hepatitis. *Lancet.* 1985; 30 March:743–745.
2. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood borne nonA, nonB viral hepatitis genome. *Science.* 1989;244:359–362.
3. Reyes GR, Purdy MA, Kim JP, et al. Isolation of a cDNA from the virus responsible for enterically transmitted nonA, nonB hepatitis. *Science.* 1990;247: 1335–1339.
4. Alter MJ, Hadler SC, Judson FN, et al. Risk factors for acute nonA, nonB hepatitis in the United States and association with hepatitis C virus infection. *JAMA.* 1990;264:2231–2235.
5. Mele A, Stazi MA, Corona R, et al. Decline of incidence of A, B and nonA, nonB hepatitis in Italy. Results of four years surveillance (1985–88). *Ital J Gastroenterol.* 1990;22:274–280.
6. Dubin N, Pasternack BS. Risk assessment for case/control subgroups by polychotomous logistic regression. *Am J Epidemiol.* 1986;123:1101–1117.
7. BMDP. *Statistical Software Manual.* Vol 2. University of California; 1990:1047–1076.
8. Mele A, Stazi MA, Gill ON, Pasquini P, SEIEVA Collaborating Group. Prevention of hepatitis B in Italy: lessons from surveillance of type-specific acute viral hepatitis. *Epidemiol Infect.* 1990;104:135–141.
9. Alter MJ, Coleman PJ, Alexander WJ, et al. Importance of heterosexual activity in the transmission of hepatitis B and nonA, nonB hepatitis. *JAMA.* 1989;262:1201–1205.
10. Pasquini P, Mele A, Stazi MA. The importance of heterosexual transmission of hepatitis B in Italy. In: Piot P, André FE, eds. *Proceedings of the International Meeting on Hepatitis B: A Sexually Transmitted Disease in Heterosexuals, Barcelona, 7 May 1990.* Amsterdam, The Netherlands: Elsevier Science Publishers; 1990:23–29.
11. Corona R, Prignano G, Mele A, et al. Heterosexual and homosexual transmission of hepatitis C virus: relation with hepatitis B virus and human immunodeficiency virus type 1. *Epidemiol Infect.* 1991: 107:667–672.

CDC's Consensus Set of Health Status Indicators: Monitoring and Prioritization by State Health Departments

Sharon L. Zucconi, PhD, and Catherine A. Carson, PhD

ABSTRACT

A survey assessed the extent to which state health departments monitor and prioritize the Centers for Disease Control and Prevention's consensus set of health status indicators. A response rate of 100% was obtained. Although mortality indicators are often monitored, only 75.5% of the states monitor work-related injury deaths. Most states monitor the incidence of acquired immunodeficiency syndrome, measles, tuberculosis, and syphilis. Low birthweight, births to adolescents, and lack of prenatal care are monitored in nearly all states and are considered high-priority problems. Only 46.9% of states are monitoring poor air quality, and only 58.8% are monitoring childhood poverty. Survey results suggest a need for standardized assessment of indicators for policy development and program planning. (*Am J Public Health.* 1994;84:1644–1646)

Introduction

In July 1991, a consensus set of 18 health status indicators was released by the Centers for Disease Control and Prevention (CDC) to assist communities in assessing their general health status.[1] These indicators addressed the requirements of the Year 2000 Objectives Planning Act[2] and objective 22.1 of *Healthy People 2000*.[3]

Priority in selecting the indicators was given to measures for which data are readily available and measures that are commonly used in public health.[1] It is not known, however, whether these indicators are actually being monitored within each state. Therefore, we conducted a survey to determine how the CDC's consensus set of health status indicators (henceforth referred to as "consensus indicators") is being assessed in the United States.

Methods

A four-page survey was mailed to directors of state health departments, as well as the District of Columbia health department, to elicit information about efforts in monitoring CDC's consensus indicators. Questions were also asked about prioritization of health problems and the sources and availability of data to monitor the indicators.

Results

An initial mailing in summer 1992, followed by two mailings to nonresponders, resulted in a cumulative response rate of 100%. As shown in Table 1, most of the consensus indicators are being monitored in all states. In general, mortality indicators are often monitored. However, only 75.5% of the states monitor work-related injury deaths. Most state health departments monitor indicators of acquired immunodeficiency syndrome (AIDS), measles, tuberculosis, and syphilis incidence, which are diseases reportable to the CDC. The incidence of low birthweight, births to adolescents, and the lack of prenatal care are being monitored in nearly all states, and these areas are considered high priorities. However, only 46.9% and 58.8% of states, respectively, are monitoring poor air quality and childhood poverty.

The most common source of mortality data was vital statistics. Two exceptions were (1) Fatal Accident Reporting System data and police reports for monitoring deaths from motor vehicle crashes and (2) worker's compensation data for occupational injuries.

Provision of mortality and incidence data to local health departments was reported by all states with local departments. However, risk factor data were often not available at the county or community level.

Discussion

This paper clearly demonstrates the widespread use of CDC's consensus set of 18 health status indicators by state health departments. However, it also highlights problems in data collection, maintenance, and resource allocation to effectively monitor these indicators.

The consensus indicators were identified, in part, because it was believed that data were readily available at national, state, and local levels.[1] The results of this survey indicate, however, that data for some indicators may actually *not* be readily available. And the availability of

At the time of the study, the authors were with the Health Policy Institute, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pa. Sharon L. Zucconi is now with the Nemours Children's Clinic, Health Services Research Center, Jacksonville, Fla. Catherine A. Carson is now with the Department of Ophthalmology, University of Melbourne, Melbourne, Australia.

Requests for reprints should be sent to Sharon L. Zucconi, PhD, Nemours Children's Clinic, Health Services Research Center, 807 Nira St, Jacksonville, FL 32207.

This paper was accepted February 1, 1994.

data had a major influence on (1) whether or not states placed a high priority on the need to monitor and (2) subsequent monitoring of indicators. States were less likely to monitor an indicator if data were unavailable, perceived of lesser quality, or of lower priority.

Sources of information used to monitor each objective may vary for some indicators (i.e., motor vehicle crash deaths, work-related injury deaths). The use of different data sources introduces the possibility of poor reproducibility of rates; differences in rates across communities may be due to different sources of data rather than actual discrepancies in events.

Differences in monitoring may be artifactual (i.e., events may be recorded but not considered indicators unless health departments actually use the information for program planning and/or resource allocation). For example, all states have *access* to mortality data, but these data may not be used for health status indicators. The disparity between data collection and use of data for program planning and evaluation is probably even greater at the local health department level. In the Institute of Medicine's view, assessment is a core function of public health.[4] Hence, the rapid falloff of monitoring beyond the vital statistics indicators and the mixed priority for even those more readily accessible data suggest areas for further research.

States were often not monitoring two consensus indicators: childhood poverty and the proportion of persons living in counties exceeding Environmental Protection Agency standards for air quality (poor air quality). One of the reasons for not monitoring poor air quality may be a lack of available data for counties (i.e., not all counties have air quality stations). It was frequently mentioned that census data are available indicating the number of children less than 18 years of age living in poverty. However, the indicator is for children less than 15 years of age living in poverty. A solution to this problem would be to either have current population survey data compiled for children less than 15 years old or change the indicator to the proportion of children less than 18 years old living in families at or below the poverty level.

Objective 22.1 of *Healthy People 2000*[3] states that the value of public health agencies in assessing the health status of the population is "enhanced when comparisons to some benchmarks, such as other similar populations or to national norms, can be made."[3(p554)] The lack of targets for some of the indicators (e.g., all-cause mortality and childhood poverty) may be a deterrent to monitoring. These indicators are not *Healthy People 2000* objectives[3]; therefore, national targets have not been established for them. In addition, several of the indicators, although similar to specific objectives, do not conform exactly (e.g., cardiovascular disease deaths); precise national targets are not available for this group of indicators. Furthermore, although definitions and national data have been recently published[5] for the consensus indicators, technical support from the CDC is needed to assist communities in establishing local targets.

As stated in objective 22.1,[3(p554)] the indicators should include measures needed for areawide health planning and for allocating health resources among programs, areas, or regions. The key to evaluating progress and comparing results across geographic, political, and programmatic lines is comparability of data. CDC's consensus indicators are widely recognized by the states and are, in effect, being monitored rather consistently, except for a few. This is a first step in achieving comparable data methods and is a key, initial step toward what the Institute of Medicine envisioned as a uniform national data set.[4] The status of these efforts should be assessed frequently as funding changes and indicators are revised.

The results of this survey are useful for evaluating the current and future impact of the consensus indicators. They also indicate a need to provide support for assessing these indicators in a standard fashion, so that they can be compared across the nation, and for policy development and program planning. □

TABLE 1—Health Departments' (n = 51) Monitoring and Prioritization of the Centers for Disease Control and Prevention's Consensus Set of Health Status Indicators

Indicator	Health Departments That Monitor, No. (%)	Departments' Priority		
		High, %	Medium, %	Low, %
Race/ethnicity-specific infant mortality[a]	49 (96.1)	81.4	14.0	4.6
Motor vehicle crash death rate[b]	50 (100)	55.8	27.9	16.3
Work-related injury death rate[b]	37 (75.5)	31.0	38.0	31.0
Suicide rate[b]	49 (98.0)	41.8	32.6	25.6
Lung cancer death rate[b]	50 (100)	72.1	25.6	2.3
Breast cancer death rate[b]	50 (100)	79.1	20.9	0
Cardiovascular disease death rate[b]	49 (98.0)	69.8	27.9	2.3
Homicide rate[b]	48 (96.0)	27.9	44.2	27.9
All-cause mortality rate[b]	46 (92.0)	35.9	43.6	20.5
AIDS incidence[c]	51 (100)	86.0	14.0	0
Measles incidence[c]	48 (94.1)	72.1	20.9	7.0
Tuberculosis incidence[c]	51 (100)	69.8	23.2	7.0
Syphilis incidence[c]	50 (98.0)	59.5	21.4	19.1
Incidence of low birthweight[d]	50 (100)	90.9	9.1	0
Births to adolescents[e]	50 (100)	90.9	9.1	0
Prenatal care[f]	48 (96.0)	97.7	2.3	0
Childhood poverty[g]	30 (58.8)	46.1	30.8	23.1
Proportion of persons living in counties exceeding Environmental Protection Agency standards for air quality[h]	23 (46.9)	20.5	35.9	43.6

Note. The denominator for percentages excludes those states with missing responses. The indicators listed are the CDC's.[1]

[a]As measured by the rate (per 1000 live births) of deaths among infants less than 1 year of age.

[b]Death rate per 100 000 population, age adjusted to the 1940 standard population.

[c]Reported incidence per 100 000 population.

[d]As measured by the percentage of the total number of live-born infants weighing less than 2500 g at birth.

[e]Births to adolescents (females aged 10–17 years) as a percentage of total live births.

[f]As measured by the percentage of mothers who delivered live infants who did not receive prenatal care during the first trimester.

[g]As measured by the proportion of children less than 15 years of age living in families at or below the poverty level.

[h]During the previous year.

Acknowledgments

Funding for this research was provided by the Health Policy Institute, University of Pittsburgh.

We thank Ransom Towsley for his helpful comments on the manuscript.

References

1. Centers for Disease Control. Consensus set of health status indicators for the general assessment of community health status—United States. *MMWR.* 1991;40:449–451.
2. Year 2000 Planning Act, Pub L No. 101-582, 42 USC § 246 (1990).
3. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services, Public Health Service; 1990. DHHS publication PHS 91-50212.
4. Institute of Medicine. *The Future of Public Health.* Washington, DC: National Academy Press; 1988.
5. Klein RJ, Hawk SA. Health status indicators: definitions and national data. *Healthy People 2000 Statistical Notes.* (Hyattsville Md: National Center for Health Statistics.) Spring 1992;1(3):1–8. DHHS publication PHS 92–1237.

Medicare-Certified Home Health Services: National and Regional Supply in the 1980s

ABSTRACT

The number of Medicare-certified home health agencies nearly doubled from 1980 to 1990. Using Health Care Financing Administration data, this study documented national and regional patterns of entry and exit by Medicare home health providers from 1980 to 1990. Nationally, agency origination rates accelerated during the early 1980s and then dropped abruptly in the second half of the decade. The proprietary sector, accounting for approximately 42% of agencies in existence during the period of the study, exhibited the greatest volatility. Regional differences are also evident. Both expansion and contraction in Medicare home health services appear to be a response to the incentives of legislation implemented during this period. (*Am J Public Health.* 1994;84:1646–1648)

Cynthia C. Scalzi, PhD, FAAN, Jacqueline S. Zinn, PhD, Michael J. Guilfoyle, and Sondra T. Perdue, DrPH

Introduction

Between 1967 and 1990, Medicare payments for home health care increased from $25 million to an estimated $4 billion, making such care the fastest growing sector of Medicare expenditures.[1–3] Although demand factors are consistent with continuing industry growth, exit from the Medicare market by home health care providers has been more common than entry since 1987.[4] This study documented variations in the national and regional distribution of Medicare-certified home health agencies during the 1980s.

Methods

Using Health Care Financing Administration (HCFA) data, we constructed a database consisting of all Medicare-certified home health agencies in existence from January 1, 1980, through June 30, 1990 (n = 8593). Agencies were grouped into one of four categories assigned by the Medicare program. A proprietary agency is owned and operated for profit by an individual or a business corporation. Government agencies are run by state, county, or city or other local municipalities. Voluntary agencies are governed by community-based boards and financed by earnings and contributions (mainly visiting nurse associations). Nonprofit agency is a residual category consisting of private, nonprofit facilities operated by all other entities. From these data, we calculated (1) the rate of home health agency origination and termination nationally between 1980 and 1990, (2) regional patterns of origination and termination in each of the 10 HCFA designated fiscal intermediary regions, and (3) variation in Medicare home health agency market entry and exit by agency category. Although identified by the city location of the HCFA regional office, each region represents several states. For example, the Boston region covers all of New England.

The origination rate, defined as the annual incidence of newly certified home health agencies, was calculated as the number of newly certified agencies in the current year divided by the number of operating agencies in the prior year. Agencies established before 1980 were included in the denominator, but not the numerator, of the calculation. The termination rate, defined as the annual rate of attrition from the Medicare program, was calculated as the number of home health agencies terminating participation in the current year divided by the peak number

Cynthia C. Scalzi is with the School of Nursing and the Wharton School, and Michael J. Guilfoyle is with the University Data Center, both at the University of Pennsylvania, Philadelphia. Jacqueline S. Zinn is with the Department of Health Administration, Temple University, Philadelphia. Sondra T. Perdue is with Analysis Plus, Federal Way, Wash.

Requests for reprints should be sent to Cynthia C. Scalzi, PhD, FAAN, School of Nursing, 342 NEB, University of Pennsylvania, Philadelphia, PA 19104-6096.

This paper was accepted March 22, 1994.

of certified agencies in operation in the current year.

Results

The origination rate for Medicare-certified home health agencies accelerated during the early 1980s and then dropped abruptly during the latter part of the decade (Figure 1). The Medicare program certified only 233 new agencies in 1980. That number peaked at 1213 in 1984 and then declined to 297 in 1988. Medicare termination rates peaked later in the decade than did origination rates. In 1980, 89 agencies lost their Medicare certification; this number increased to 523 in 1987 and dropped to 263 in 1989. Six regions (Atlanta, Chicago, Denver, Kansas City, Philadelphia, and Seattle) exhibit patterns of market entry across time that approximate the national experience. In Dallas and San Francisco, market entry peaked slightly earlier, in 1983, but exhibited the same pattern of decline and leveling through the remainder of the decade. By contrast, the patterns of origination over time in the Boston and New York regions are relatively flat. All regions approximate the national pattern of terminations portrayed in Figure 1, with rates peaking in 1987 or 1988.

Of the 8593 certified agencies in operation at any time from 1980 to mid-1990, 1508 (17.6%) were voluntary, 2007 (23.4%) were governmental, 3620 (42.1%) were proprietary, and 1458 (17.0%) were private nonprofit. Nationally, rates of origination and termination vary by agency category. Of the 1508 voluntary agencies in operation at any time during the decade, 659 (43.9%) originated and 333 (22.1%) terminated. In comparison, 3284 (90.7%) of the 3620 proprietary agencies originated, and 1551 (42.9%) terminated, during the decade. Originations among private nonprofit home health agencies occurred almost as rapidly: 1078 (73.9%) during the decade. Terminations among private nonprofit agencies were less frequent. Of the 1458 existing, 371 (25.5%) terminated during the decade. For the 2007 government agencies, origination and termination rates were essentially the same: 29.4% (591) and 31.7% (637), respectively.

Seven regions had net percentage increases in the number of home health agencies in operation between 1979 and 1990 that exceeded the national average of 99% (Table 1). The vast majority of agencies in all regions in 1979 were classified as either voluntary or governmental (Table 2). As indicated in Table 3, originations by agency type of control showed substantial regional effects. In virtually all regions, the agency category with the greatest market share at the beginning of the decade also accounted for the highest proportion of new entrants during the decade. In seven of the regional markets, the highest proportion of terminations occurred in the proprietary category (Table 3). In every region, with the exception of Kansas City, the control type with the greatest percentage of originations during the decade also experienced the greatest percentage of terminations.

It is clear from the difference in the rates of origination and termination that some regional markets were more volatile than others. For example, 70% of San Francisco region proprietary agencies in existence at any time during the decade terminated Medicare participation (Table 2). Some regions, notably Kansas City and New York, showed very little change in market share by agency type from the beginning to the end of the decade.

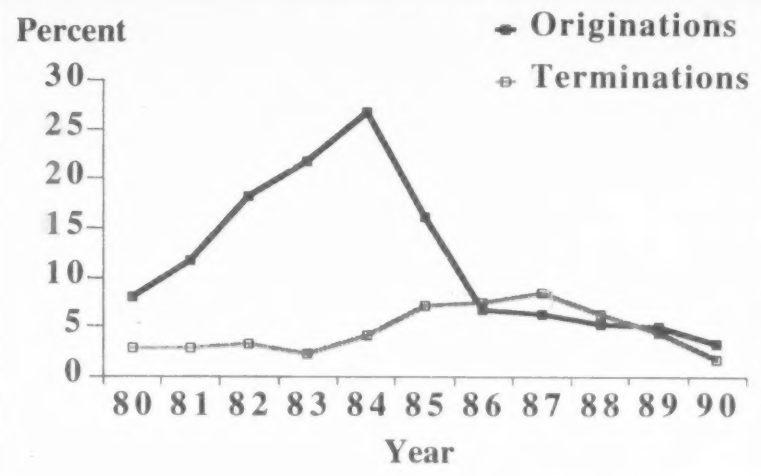


FIGURE 1—Annual rates of origination and termination of Medicare-certified home health agencies in the United States, 1980 to 1989.

TABLE 1—Regional Patterns of Home Health Care Agency Originations, Terminations, and Net Change, 1979 to 1990

	Total of Region as of 1979	Total Agency Originations, 1980–1990	Total Agency Terminations, 1980–1990	Total Agencies as of 1990	Net Increase, 1979–1990, %
All agencies	2935	5747	2847	5836	99
Atlanta	721	1050	379	1392	93
Boston	301	345	113	533	77
Chicago	519	1128	449	1198	131
Dallas	336	951	875	412	23
Denver	120	268	145	243	103
Kansas City	231	510	186	555	140
New York	200	301	38	463	132
Philadelphia	250	579	287	542	117
San Francisco	176	542	308	410	133
Seattle	81	154	66	169	109

Discussion

Industry growth early in the decade may be linked to legislation that deregulated the industry and stimulated privatization. With the passage of the Omnibus Budget Reconciliation Acts of 1980 and 1981, state licensure requirements for

TABLE 2—Distribution of Home Health Care Agencies, by Type of Control, 1980 to 1990

	No. (Market Share, %)									
	Total Agencies as of 1980					Total Agencies as of 1990				
Region	Proprietary	Nonprofit	Voluntary	Government	Total	Proprietary	Nonprofit	Voluntary	Government	Total
Atlanta	164 (23)	116 (16)	75 (10)	366 (51)	721	460 (33)	270 (19)	122 (9)	540 (39)	1392
Boston	0 (0)	15 (5)	233 (77)	53 (18)	301	47 (9)	52 (10)	375 (70)	59 (11)	533
Chicago	26 (5)	74 (14)	140 (27)	279 (54)	519	119 (10)	236 (20)	325 (27)	518 (43)	1198
Dallas	52 (15)	46 (14)	58 (17)	180 (54)	336	10 (2)	90 (22)	78 (19)	234 (57)	412
Denver	3 (3)	12 (10)	27 (23)	78 (65)	120	21 (9)	65 (27)	57 (23)	100 (41)	243
Kansas City	1 (0)	26 (11)	41 (18)	163 (71)	231	12 (2)	79 (14)	119 (21)	345 (62)	555
New York	1 (1)	30 (15)	85 (43)	84 (42)	200	22 (5)	83 (18)	201 (43)	157 (34)	463
Philadelphia	18 (7)	41 (16)	107 (43)	84 (34)	250	79 (15)	129 (24)	227 (42)	107 (20)	542
San Francisco	60 (34)	6 (3)	57 (32)	53 (30)	176	65 (16)	46 (11)	172 (42)	127 (31)	410
Seattle	11 (14)	14 (17)	26 (32)	30 (37)	81	19 (11)	46 (27)	58 (34)	46 (27)	169
Total	336 (11)	380 (13)	849 (29)	1370 (47)	2935	854 (15)	1096 (19)	1690 (29)	2195 (38)	5836

TABLE 3—Distribution of Home Health Care Agency Terminations and Originations, by Type of Control, 1980 to 1990

	No. (% of All Originations or Terminations)									
	Originations					Terminations				
Region	Proprietary	Nonprofit	Voluntary	Government	Total	Proprietary	Nonprofit	Voluntary	Government	Total
Atlanta	460 (44)	205 (20)	85 (8)	300 (29)	1050	164 (43)	51 (13)	38 (10)	126 (33)	379
Boston	75 (22)	41 (12)	199 (58)	30 (9)	345	28 (24)	4 (4)	57 (50)	24 (21)	113
Chicago	352 (31)	227 (20)	238 (21)	311 (28)	1128	259 (58)	65 (14)	53 (12)	72 (16)	449
Dallas	488 (51)	169 (18)	78 (8)	216 (23)	951	530 (61)	125 (14)	58 (7)	162 (19)	875
Denver	68 (25)	73 (27)	48 (18)	79 (29)	268	50 (34)	20 (14)	18 (12)	57 (39)	145
Kansas City	107 (21)	80 (16)	97 (19)	226 (44)	510	96 (52)	27 (15)	19 (10)	44 (24)	186
New York	28 (9)	59 (20)	134 (45)	80 (27)	301	7 (18)	6 (16)	18 (47)	7 (18)	38
Philadelphia	229 (40)	130 (22)	150 (26)	70 (12)	579	168 (59)	42 (15)	30 (10)	47 (16)	287
San Francisco	220 (41)	65 (12)	150 (27)	107 (20)	542	215 (70)	25 (8)	35 (11)	33 (11)	308
Seattle	42 (27)	38 (25)	39 (25)	35 (23)	154	34 (52)	6 (9)	7 (11)	19 (29)	66
Total	2069 (36)	1087 (19)	1175 (20)	1416 (25)	5747	1551 (54)	371 (13)	334 (12)	591 (21)	2847

Note. Originations and terminations occurring prior to 1980 are not included.

Medicare certification were gradually eliminated, opening a floodgate for the entry of proprietary agencies in the home care field.[4,5] The Tax Equity and Fiscal Responsibility Act of 1982 introduced prospective payment for hospitals while retaining cost-based retrospective reimbursement for home care, contributing to the proliferation of hospital-sponsored agencies.[1,7] Paradoxically, whereas for-profit agencies aggressively entered markets in virtually all regions, there was also a higher percentage of terminations among such agencies, averaging 43% nationally as compared with 22% for voluntary agencies and 26% for private nonprofit agencies. Lack of policy coordination may well have contributed to these patterns of growth and decline. In 1985 and 1986, fiscal intermediaries took measures that placed more restrictive conditions on eligibility for home health services under Medicare. As a result, increasing numbers of claims were retrospectively denied on the basis of lack of medical necessity or procedural technicalities.[7] Thus, while some federal policies were intended to stimulate private sector growth and competition, the reimbursement policies of Medicare intermediaries may have thwarted that intent, leading to an increasing termination of participation that peaked in 1987. Further research is needed to assess the impact of growth and decline in Medicare home health participation on access in selected markets. □

References

1. Wood JB. The effects of cost-containment on home health agencies. *Home Health Services Q.* 1986;10(4):59–78.
2. *Health Care Financing: Program Statistics.* Washington, DC: Health Care Financing Administration; 1989. HFCA publication 03270.
3. Rappaport M, Wood JB. Financing and home health agencies. *Home Health Services Q.* 1989;10(3/4):131–147.
4. Scalzi C, Meyer M. Model of factors impacting closures and mergers of VNA home care agencies. *Home Health Services Q.* 1992;12(4):113–124.
5. Omnibus Budget Reconciliation Act of 1980. Pub L. No. 96-499.
6. Omnibus Budget Reconciliation Act of 1981. Pub L. No. 97-35.
7. Keenan JM, Fanale JE, Ripsin C, Billows L. A review of federal home-care legislation. *J Am Geriatr. Soc.* 1990;38:1041–1048.

Gynecologists' Sex, Clinical Beliefs, and Hysterectomy Rates

Nina A. Bickell, MD, MPH, Jo Anne Earp, ScD, Joanne M. Garrett, PhD, and Arthur T. Evans, MD, MPH

ABSTRACT

This study determined how gynecologists' sex, beliefs, appropriateness ratings, and practice characteristics influence hysterectomy rates in North Carolina. Gynecologists who performed hysterectomies at higher rates were further from training, practiced in areas with fewer gynecologists, and had more patients with abnormal bleeding or cancer. Male gynecologists performed 60% more hysterectomies than female gynecologists, but this may have been because they were further from their training. Appropriateness ratings were affected by gynecologists' attitudes toward surgery, recency of training, and practice case mix, and by patients' expressed desire to avoid surgery, but they did not predict hysterectomy rates. To decrease their chances of undergoing hysterectomy, patients should express their preferences and possibly seek the opinion of more recently trained gynecologists. (*Am J Public Health.* 1994;84: 1649–1652)

Introduction

Tremendous variation exists in rates of hysterectomy, the second most common major surgery performed in the United States.[1] Patient demographics and use of health care resources have not explained much of this variation.[2] Some of it is accounted for by physician age and sex, but most has been attributed to physician uncertainty about the appropriate indications for hysterectomy.[2–5] Uncertainty results from limited information about hysterectomy outcomes, physician judgment, and patient preferences for treatment.[3] Unstudied is the relationship between hysterectomy rates and practicing physicians' beliefs about the appropriateness of surgery or the potential impact of patient preferences. Given the increasing number of female gynecologists,[6] patients' preference for female gynecologists,[7] and the wide variation in hysterectomy rates, we surveyed surgically active gynecologists in North Carolina, a state with one of the highest hysterectomy rates, and studied the association between rates and physicians' sex, ratings of hysterectomy's appropriateness, and beliefs about surgery and the uterus.

Methods

Study Population and Data Collection

Surgically active gynecologists who had completed residency training in gynecology and were licensed by the North Carolina Board of Medical Examiners (n = 671) were eligible for the study. A 106-question survey was mailed in 1990 to all 77 female and 154 randomly selected male gynecologists. The survey included questions about attitudes toward surgery and the uterus; practice characteristics; demographics; and clinical scenarios of women presenting with cervical dysplasia (the most common equivocal indication), bleeding fibroids and/or painful fibroids (the most prevalent indication for hysterectomy), and fear of cancer (a common subjective indication).[2] A follow-up survey to assess case mix was mailed to the 140 respondents of the initial survey.

Statistical Analyses

The hysterectomy rate was calculated using the average number of hysterectomies per month, divided by the number of patients seen per month, multiplied by 100. Gynecologists' self-reported hysterectomy rates were correlated with rates from the North Carolina Hospital Discharge Database ($r = .99$) to confirm their validity. Bivariate and multivariable relationships between hysterectomy rate, attitudes and practice characteristics, and physician demographics were assessed. Physician density by county was calculated by dividing the number of gynecologists in the county by the number of women older than age 15 years in that county.

All scenarios asked about the appropriateness of a hysterectomy for a 45-year-old woman, approximately the average age when women undergo that procedure. Appropriateness was measured on a 9-point scale, on which 1 was extremely inappropriate and 9 was extremely appropriate.

The impact of patients' desire to avoid hysterectomy was assessed by having physicians rate each scenario's appropriateness for women who did and did *not* express an aversion to the procedure. *T* tests compared the mean difference in appropriateness scores.

Factor analysis confirmed physicians' attitudes toward (1) the uterus' contribution to sexual function (six items, Cronbach's alpha = .84); (2) the importance of the uterus for sexual identification (four items, Cronbach's alpha = .92); and (3)

At the time of this study, Nina A. Bickell and Arthur T. Evans were with the Department of Medicine at the University of North Carolina, Chapel Hill. Jo Anne Earp is with the Department of Health Behavior and Health Education, School of Public Health, and Joanne M. Garrett and Dr Evans are with the Cecil Sheps Center for Health Services Research, all at the University of North Carolina, Chapel Hill.

Requests for reprints should be sent to Nina A. Bickell, MD, MPH, New York State Department of Health, Office of Quality Improvement, Corning Tower, Room 1223, Albany, NY 12237.

This article was accepted March 16, 1994.

TABLE 1—Physician and Practice Characteristics

	Male (n = 79)	Female (n = 28)	P[a]
Physician demographics			
Age, mean y (range)	48.7 (33–72)	38.9 (31–75)	< .0001
Year of residency graduation (range)	1973 (1950–1987)	1983 (1940–1989)	< .0001
Physician attitudes[b]			
Hysterectomy more efficacious than medical treatment: 6-pt scale, mean (SD)	3.4 (0.7)	3.9 (0.8)	.02
Uterus contributes to sexual function: 6-pt scale, mean (SD)	4.1 (1.1)	3.9 (1.3)	NS[c]
Uterus contributes to gender identification: 4-pt scale, mean (SD)	2.4 (0.6)	2.4 (0.7)	NS
Practice characteristics			
Private insurance, %	70	73	NS
< 400-bed hospital, %	62	50	NS
Community practice, %	96	92	NS
University practice, %	4	8	NS
Patients < 45 years old, % (range)	55 (10–80)	56 (30–90)	NS
Physician density (no. females > 15 years/ no. ob-gyns)	3646	2707	.02
Case mix			
Abnormal bleeding, mean (range)	17.4 (2–70)	15.7 (1–30)	NS
Pelvic pain, mean (range)	16.3 (0–70)	13.2 (1–38)	NS
Pelvic relaxation, mean (range)	6.7 (0.5–50)	4.9 (0–20)	NS
Cancer, mean (range)	2.4 (0–75)	1.3 (0–5)	NS
Dysplasia, mean (range)	8.3 (0–30)	13.6 (1–60)	NS

[a]Comparing male with female gynecologists.
[b]Controlling for year of residency training.
[c]NS = not significant.

the relative efficacy of hysterectomy compared with medical treatment (six items, Cronbach's alpha = .66) (see Appendix).

Case mix was calculated from physicians' reports of the percentage of all patients who had intact uteri and signs or symptoms of abnormal bleeding, pelvic pain, pelvic relaxation, dysplasia, or cancer.

Response Rate

For the initial survey, 190 of 231 gynecologists were surgically active and eligible for inclusion in the study; the overall response rate was 74% (140/190). Of the initial survey respondents who were still alive and practicing at the time of second survey, 92% (107/116) responded to the follow-up. This represents 50% (28/56) of the eligible women and 59% (79/134) of the men of the original sample.

Results

Sample Characteristics

The mean rate of hysterectomies performed for the entire sample was 1.3 per month (SD = 1.3; range: 0, 6.7), with, on average, 1.4 performed by male gynecologists compared with 0.9 performed by female gynecologists (risk ratio [male/female] = 1.6; $P = .001$). Differences in other characteristics by sex are presented in Table 1.

Factors Associated with Hysterectomy Rates

Gynecologists who tended to perform hysterectomies at higher rates were male ($P = .001$), were further from training ($r = .26$; $P = .007$), practiced in areas with a lower density of gynecologists ($r = .25$; $P = .01$), and had a higher proportion of younger patients ($r = .26$; $P = .007$) and patients with abnormal bleeding ($r = .31$; $P = .001$) or cancer ($r = .25$; $P = .008$). In the multivariable model, the only physician factor that remained as a predictor of hysterectomy rate was year of residency graduation. Physicians' sex appears to be related to hysterectomy rates, which may be explained by differences in recency of training. Insurance reimbursement and type of practice are not predictive of hysterectomy rates. Recently trained gynecologists perform hysterectomies at lower rates than gynecologists who have been in practice longer, are more likely to believe that the uterus contributes to sexual function ($r = .26$; $P = .002$), and are less likely to believe that hysterectomy is generally more efficacious than medical management ($r = .33$; $P = .0002$). Physicians who practice in areas with a high density of gynecologists trained more recently ($r = .28$; $P = .004$) and are less inclined to believe that surgery is a better treatment than medicine ($r = .28$; $P = .004$).

Factors Associated with Appropriateness of Hysterectomy Ratings

Beliefs about the appropriateness of hysterectomy for the various clinical scenarios are related to several physician factors but are not related to hysterectomy rates (Table 2). Beliefs about the relative efficacy of hysterectomy are related to appropriateness ratings, but these beliefs are also not associated with higher hysterectomy rates.

Impact of Patient Preference

When patients expressed a desire to avoid surgery, physicians' ratings of the appropriateness of hysterectomy declined for all seven scenarios. This decrease was particularly prominent for scenarios concerning painful fibroids (mean difference = 3.1; $P = .001$), bleeding fibroids responding to treatment (mean difference = 2.3; $P = .001$), and cervical dysplasia (mean difference = 2.3; $P = .001$). Male and female gynecologists were similarly affected by a patient's expressed preference to avoid surgery.

Discussion

Some researchers have implicated a "physician factor" as a source of practice variation.[5,8–10] This may represent several sets of beliefs, which, in concert, may exert a more powerful effect on behavior than they do individually. Year of resi-

dency training, a major influence on physician practice style,[11] may serve as a proxy for numerous beliefs held by different cohorts of trainees about the uterus' contribution to sexual function and sexual identification, about surgery, and perhaps about other issues we did not measure. As noted, recently trained gynecologists have lower rates of hysterectomy, are less inclined to believe that hysterectomy is superior to medical management, and are more likely to believe that the uterus contributes to a woman's sexual functioning.

Socialization during training affects the formation and acquisition of professional beliefs and practices.[12] Within similar residency-year cohorts, female and male gynecologists did not differ in rates or ratings of hysterectomy, or in attitudes about the "uterus' contribution to a woman." The increasing number of women in obstetrics-gynecology may influence existing norms and account for our inability to detect sex differences. It is possible that women gynecologists have adopted their contemporaries' "male medical viewpoint" as the majority of gynecologists trained during these years are men.

Two practice factors are also important predictors of high hysterectomy rates: practice case mix and geographic location (i.e., physician-to-population density). Gynecologists with a higher percentage of patients with cancer or abnormal bleeding perform hysterectomies at higher rates. Although this correlation may simply reflect the increased need for hysterectomy among patients with significant symptoms or potentially dangerous diagnoses, it may also reflect a lower threshold for performing surgery among physicians exposed to more severe pathology.

We found that physicians who work in areas with fewer gynecologists perform hysterectomies at higher rates, and believe that surgery is a better treatment than medicine. These findings may reflect shared professional beliefs and community norms in specific geographic areas. Greater numbers of surgeons are not necessarily associated with higher rates of surgery.[9] Our study supports this. It is not necessarily the oversupply of physicians that affects hysterectomy rates; more likely, it is physician beliefs and attitudes, as these reflect community norms, that help determine variations in practice patterns.

Gynecologists' beliefs about the appropriateness of hysterectomy do not predict the rates at which hysterectomies are performed. This lack of association is consistent with other studies that have not found an association between appropriateness ratings and rates of cardiac angiograms, carotid endarterectomy, or gastrointestinal endoscopy.[13] This disjunction between ratings and rates provides support for particular physician factors (unmeasured in this study) rather than simply the appropriateness of a procedure influencing practice.

Greater satisfaction is found between female patients and female physicians than among other types of physician-patient dyads.[14,15] Researchers have found female physicians to be more sensitive than male physicians to issues of appropriate health care, contraception, hormone replacement therapy, abortion, and sex discrimination.[16–18] Some have speculated that increasing the number of women in medicine will make medicine more humane, particularly in the area of reproductive health.[19] It is somewhat surprising, then, that we did not find a difference by sex in physicians' ratings of appropriateness for the clinical scenarios, especially those in which women expressed a preference for avoiding surgery.

Patient preferences, doctor-patient communication, and physicians' beliefs about the efficacy of surgery are important factors in making treatment decisions for elective procedures.[2,20] Of all the factors we examined, a woman's expressed desire to avoid surgery had the greatest impact on lowering the appropriateness score. Social desirability in response to a clinical simulation may provide part of the explanation. Although appropriateness scores do not correlate with hysterectomy rates in this study, patients' expressed preferences could influence surgical referral. Patients concerned about undergoing possibly unnecessary surgery may want to seek opinions of gynecologists who have trained more recently. Perhaps even more important, patients should be encouraged to be open about their concerns and preferences for, or against, surgery when consulting with their physicians. □

TABLE 2—Factors Associated with Hysterectomy Ratings

Scenario	Mean Score (SD)	Associated Factors[a]	*P*
Cervical punch biopsy with dysplasia	4.8 (3.0)	. . .	NS[b]
Cervical cone biopsy with dysplasia	7.3 (2.2)	. . .	NS
Cancer fear in a woman with cancer risk factors	4.1 (2.5)	Residency year, attitude toward surgery, patients < 45 years old, MD sex	.0001
Bleeding fibroid refractory to iron treatment	8.6 (0.9)	Attitude toward surgery	NS
Bleeding fibroid responsive to iron treatment	6.5 (2.0)	Attitude toward surgery, patients with abnormal bleeding or cancer, MD sex	.001
Painful fibroid in otherwise healthy woman	7.5 (1.7)	Attitude toward surgery, patients with cancer, patients < 45 years old	.07
Painful fibroid refractory to medical treatment	8.3 (1.2)	Attitude toward surgery	.009

Note. The appropriateness scale for hysterectomy patients ranged from 1 to 9, with 1 = extremely inappropriate and 9 = extremely appropriate.

[a]Patients with abnormal bleeding = percentage of patients in physician's practice with abnormal bleeding; patients with cancer = percentage of patients in physician's practice with cancer. Based on multivariate analysis, linear model: appropriateness score of clinical scenario = % patients in practice with abnormal bleeding, % patients in practice with cancer, attitude toward surgery, population/physician density, physician sex, year of residency training, % practice < 45 years old.

[b]NS = not significant.

Acknowledgments

The authors gratefully thank Drs Martha Gerrity and Robert DeVellis for their help with questionnaire and scale development, Dr Wesley Fowler for his help enlisting support, and Ms Carol Porter for help with statistical programming. Dr Bickell was Robert Wood Johnson Clinical Scholar at the time of this study.

References

1. Graves EJ. National hospital discharge survey: annual summary, 1990. *Vital Health Stat [13].* 1992;112. DHHS publication PHS 92-1773

2. DeFriese GH, Evans AT, Ricketts TC, Cromartie EP. *Practice Variation Study of Hysterectomy. Final Report.* North Carolina Medical Society, Practice Pattern Variation Committee; September 1989.
3. Carlson KJ, Nichols DH, Schiff I. Indications for hysterectomy. *N Engl J Med.* 1993;328:856–860.
4. Domenighetti G, Luraschi P. Hysterectomy and sex of the gynecologist. *N Engl J Med.* December 5, 1985;313:1482. Letter.
5. Roos NP. Hysterectomy: variations in rates across small areas and across physicians' practices. *Am J Public Health.* 1984;74:327–335.
6. Page L. Will women become the new ob-gyn majority? *AMA News.* December 9, 1991; 15–16.
7. Kelly JM. Sex preference in patient selection of a family physician. *J Fam Pract.* 1980;11:427–430.
8. Leape LL, Park RE, Solomon DH, Chassin MR, Kosecoff J, Brook RH. Relation between surgeons' practice volumes and geographic variation in the rate of carotid endarterectomy. *N Engl J Med.* 1989;321:653–657.
9. Goyert GL, Bottoms SF, Treadwell MC, Nehra PC. The physician factor in cesarean birth rates. *N Engl J Med.* 1989;320:706–709.
10. Brook RH, Park RE, Chassin MR, Solomon DH, Keesy J, Kosecoff J. Predicting the appropriate use of carotid endarterectomy, upper gastrointestinal endoscopy, and coronary angiography. *N Engl J Med.* 1990;323:1173–1177.
11. Carter RE. A comparison of the emphasis given to nine therapies in residency training and later practice of psychiatrists. *Acad Med.* 1990;65:649–651.
12. Earp JA, Fletcher SW, O'Malley MS, Fletcher RH. Attitudes of internal medicine subspecialty fellows toward primary care. *Arch Intern Med.* 1984;144:329–333.
13. Chassin MR, Kosecoff J, Park RE, et al. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *JAMA.* 1987;258:2533–2537.
14. Greer S, Dickerson V, Schneiderman LJ, et al. Responses of male and female physicians to medical complaints and male and female patients. *J Fam Pract.* 1986;23:43–53.
15. Comstock IM, Hooper EM, Goodwin JM, et al. Physician behaviors that correlate with patient satisfaction. *J Med Educ.* 1982;57:105–112.
16. Weisman CS, Teitelbaum MA. Physician gender and the physician-patient relationship. Recent evidence and relevant questions. *Soc Sci Med.* 1985;20:1119–1127.
17. Margolis AJ, Greenwood S, Heilbron D. Survey of men and women residents entering United States obstetric and gynecology programs in 1981. *Am J Obstet Gynecol.* 1983;146:541–546.
18. Greendale GA, Carlson KJ, Schiff I. Estrogen and progestin therapy to prevent osteoporosis: attitudes and practices of general internists and gynecologists. *J Gen Intern Med.* 1990;5:464–469.
19. Reiman AS. Here come the women. *N Engl J Med.* 1980;302:1252–1253.
20. Barry MJ, Mulley AG, Fowler FJ, Wennberg JW. Watchful waiting vs immediate transurethral resection for symptomatic prostatism. The importance of patients' preferences. *JAMA.* 1988;259:3010–3017.

APPENDIX—Physicians' Beliefs Concerning Hysterectomy, Surgery, and the Importance of the Uterus

Importance of the uterus for sexual identification: Cronbach's alpha = 0.92
The uterus is important in making women feel
- feminine
- like other women
- normal
- self-esteem
- self-worth

Importance of uterus for sexual function: Cronbach's alpha = 0.84
The uterus is important in contributing to
- sexual functioning
- sensations during orgasm
- sexual responsiveness
- more health benefits than risks

Attitude toward hysterectomy and surgery: Cronbach's alpha = 0.66
- After hysterectomy, sex is more pleasurable for women.
- Hysterectomy is the best cure for pelvic pain.
- Pelvic pain is usually relieved with hysterectomy.
- Medicine is palliative; surgery is definitive.
- Surgery is more successful than medicine for relieving uterine pain.
- After hysterectomy, sex is more pleasurable for women's sexual partners.

Hysterectomy Use: The Correspondence between Self-Reports and Hospital Records

Kate M. Brett, PhD, and Jennifer H. Madans, PhD

ABSTRACT

Studies of the relationship between hysterectomy use and sociodemographic factors tend to use self-reported data. In the current study, data were collected from a representative sample of US women who have been prospectively followed since 1971. Hysterectomy status was obtained by self-report and from hospital records. Although these two measures of hysterectomy were highly related, more women reported hysterectomy than could be confirmed by hospital records. The two measures showed similar associations between several obstetric and demographic characteristics and hysterectomy status, suggesting that the use of self-reported hysterectomy data does not bias analyses of potentially associated factors. (*Am J Public Health.* 1994;84:1653–1655)

Introduction

Hysterectomy is the most commonly performed nonobstetric inpatient surgical procedure among women in the United States.[1] Several studies have attempted to identify factors associated with a greater likelihood of receiving this surgery in order to explain the variation in hysterectomy rates among industrialized countries.[2–8] The results have found hysterectomy to be associated with increasing age, lower education and family incomes, higher parity, and previous miscarriages. With the exception of one British study,[7] these studies have used self-reported hysterectomy information and could be affected by differential recall biases.

The purpose of this paper is to estimate the bias associated with self-report in analyses of hysterectomy. Using data from a prospective cohort study representative of the US population, we examined the correspondence between self-reported hysterectomy and hysterectomy confirmed by hospital records. We used hysterectomy rates and the association between hysterectomy and a number of socioeconomic and obstetric factors to estimate the potential bias introduced by the use of self-report.

Methods

Data were taken from the Epidemiologic Followup Study to the First National Health and Nutrition Examination Survey, a prospective investigation of 14 407 subjects 25 to 74 years of age who completed the First National Health and Nutrition Examination Survey (NHANES I) between 1971 and 1975. Follow-ups have been conducted from 1982 through 1984, in 1986, and in 1987.[9–11] We limited our analysis to Black or White women who, at the baseline survey, were between 25 and 34 years of age and reported having their womb. Women in this age group are unlikely to have had prior hysterectomy.[12] Of the 2104 Black and White women in the targeted age group, 2037 had intact uteri at baseline.

During the follow-up interviews, respondents were asked whether they had their womb and were requested to recount all hospital stays. Women were classified as reporting a hysterectomy if they reported the removal of their womb or if uterine cancer, hysterectomy (partial or full), or uterine removal was named as one of the causes of a hospitalization.

All hospitals named in the interview were contacted, if consent had been obtained, to obtain discharge summaries for all stays occurring during the interview period. Women for whom a hospital discharge summary was obtained listing hysterectomy in the procedures section (*International Classification of Diseases*, 9th edition, codes 68.3 through 68.7) were characterized as having a hospital-confirmed hysterectomy.

Agreement between the two hysterectomy measures was summarized by the kappa statistic.[13] Chi-square tests or *t* tests were used to compare the demographic and obstetric characteristics of women by hysterectomy status. Finally, proportional hazards models of hysterectomy involving each measure were estimated; these models incorporated covariates identified in previous research while accounting for differing lengths of follow-up among respondents.[14]

Results

Four hundred fifty-two respondents reported having had a hysterectomy, while 301 women had hospital-confirmed hysterectomies. Hysterectomy was coded on the hospital records of 3 women who did not report a hysterectomy. Conversely, 154 women reported having had a hysterectomy that could not be confirmed,

The authors are with the Office of Analysis, Epidemiology, and Health Promotion, National Center for Health Statistics, Hyattsville, Md.

Requests for reprints should be sent to Kate M. Brett, PhD, NCHS/OAEHP/DE, 6525 Belcrest Rd, Rm 730, Hyattsville, MD 20782.

This paper was accepted April 20, 1994.

TABLE 1—Differences in Average Characteristics of Women with Different Levels of Hysterectomy-Reporting Completeness

	Hospital-Confirmed Hysterectomy (n = 301)	Reported Hysterectomy (n = 452)	No Hysterectomy (n = 1582)
Age at baseline, y	29.9	29.9	29.1**
Age at hysterectomy, y	38.4	37.5*	
Live births, no.	3.0	3.0	2.8
Miscarriages, no.	0.6	0.6	0.4**
Body mass index, cm/g^2	24.4	24.2	24.0
Black, %	14.6	16.6	14.2
Education level, %			
Less than high school	33	32	22**
High school graduate	54	52	46
At least some college	13	16	31
Residence, %			
Central city	25.3	28.8	32.4
Suburb	32.2	31.0	32.0
Rural	42.5	40.3	35.6

*$P < .01$ (compared with hospital-confirmed hysterectomies).
**$P < .01$ (compared with all reported hysterectomies).

TABLE 2—Predictive Model of Receipt of Hysterectomy Using Two Different Methods of Outcome Classification

	Relative Risk (95% Confidence Interval)	
	Hospital-Confirmed Hysterectomy	Reported Hysterectomy
Age at baseline (per year)	1.07 (1.03, 1.12)	1.09 (1.05, 1.12)
Education level		
Less than high school	1.20 (0.92, 1.55)	1.16 (0.93, 1.43)
High school graduate	1.0[a]	1.0[a]
At least some college	0.41 (0.29, 0.59)	0.51 (0.39, 0.66)
Race		
White	1.0[a]	1.0[a]
Black	0.92 (0.65, 1.31)	1.07 (0.82, 1.40)
Residence		
Central city	1.0[a]	1.0[a]
Suburb	1.36 (0.99, 1.86)	1.14 (0.90, 1.44)
Rural	1.37 (1.02, 1.84)	1.18 (0.93, 1.49)
Parity[b]		
0–1	1.0[a]	1.0[a]
2–3	1.03 (0.73, 1.46)	1.12 (0.84, 1.48)
4+	0.97 (0.66, 1.43)	0.98 (0.72, 1.34)
Miscarriages[c]		
0–1	1.0[a]	1.0[a]
2+	1.24 (0.86, 1.80)	1.28 (0.95, 1.73)
Body mass index[d]		
< 28	1.0[a]	1.0[a]
28+	0.97 (0.73, 1.30)	0.94 (0.74, 1.20)

[a]Referent group.
[b]Data were missing for 193 subjects.
[c]Data were missing for 172 subjects.
[d]Weight in kilograms divided by height in square meters.

although validation was not possible for 33 of these subjects because of problems in obtaining hospital records.

Comparisons between reported hysterectomies and hospital-confirmed hysterectomies yielded a kappa statistic of 0.75, which signifies fairly good agreement. However, because hospital records were obtained only if a hospital stay was reported, the kappa statistic primarily reflects the agreement between not reporting a hysterectomy and failure to obtain a hospital record of hysterectomy. Conversely, only 298 (66%) of the 452 reported hysterectomies could be confirmed.

Women reported their age at hysterectomy to be about 1 year younger than indicated by hospital records (Table 1). No other significant differences between the two groups of women categorized as having a hysterectomy were found. Compared with women who had either reported or hospital-confirmed hysterectomies, women who did not have a hysterectomy were younger at baseline, had fewer miscarriages, and were more likely to have had at least some college experience.

Multivariate models that used self-reported or hospital-confirmed cases as the outcome variable yielded very similar results (Table 2). Only age and high education were significantly associated with hysterectomy by both classification methods, while urban area of residence was related to hospital-confirmed, but not self-reported, hysterectomy.

Discussion

Most investigations of the relationship between hysterectomy use and demographic and obstetric characteristics have relied on self-reported information.[2,6,8] This study was able to incorporate information obtained through both self-report and hospital records. Twenty-five percent of women in the cohort reported ever having had a hysterectomy, while only 18% of the cohort had either a hospital-confirmed hysterectomy during follow-up or a self-reported hysterectomy prior to the baseline study. Estimates of rates based on hospital records in this study probably underestimate the true rates, since obtaining records was dependent on the respondents' ability to remember hospitalization information correctly, on obtaining consent for the record review, on securing the correct address for the hospitals, and on procuring assistance from hospital personnel in locating and summarizing all hospital stays.

Regardless of the source of information, women who had had hysterectomies were found to be older at baseline and to have less formal education than women with intact uteri. Only one variable, location of residence, was found to have a different relationship to hysterectomy by data source. Given the number of variables investigated, this difference is slight.

The similarity of analysis results may not apply to responses from women older than those included in this study or when the recall period is longer. We conclude from these data that associations obtained with self-reported hysterectomy information are not unduly biased. Although high reliability of hysterectomy self-reporting had been found previously,[15] the validity had not been established. Given these findings, it appears that investigations of the sociobehavioral characteristics associated with hysterectomy use will obtain valid results using either data collection method. □

Acknowledgment

This paper was presented, in part, at the 121st Annual Meeting of the American Public Health Association, October 1993, San Francisco, Calif.

References

1. *Health, United States, 1992.* Hyattsville, Md: National Center for Health Statistics, 1993.
2. Schofield MJ, Hennrikus DJ, Redman S, Sanson-Fisher RW. Prevalence and characteristics of women who have had a hysterectomy in a community survey. *Aust NZ J Obstet Gynaecol.* 1991;31:153–158.
3. Koepsell TD, Weiss NS, Thompson DJ, Martin DP. Prevalence of prior hysterectomy in the Seattle-Tacoma area. *Am J Public Health.* 1980;70:40–47.
4. Meilahn EN, Matthews KA, Egeland G, Kelsey SF. Characteristics of women with hysterectomy. *Maturitas.* 1989;11:319–329.
5. Kjerulff K, Langenberg P, Guzinski G. The socioeconomic correlates of hysterectomies in the United States. *Am J Public Health.* 1993;83:106–108.
6. Coulter A, McPherson K. Socioeconomic variations in the use of common surgical operations. *BMJ.* 1985;291:183–187.
7. Vessey MP, Villard-Mackintosh L, McPherson K, Coulter A, Yeates D. The epidemiology of hysterectomy: findings in a large cohort study. *Br J Obstet Gynaecol.* 1992;99:402–407.
8. Santow G, Bracher M. Correlates of hysterectomy in Australia. *Soc Sci Med.* 1992;34:929–942.
9. Cohen BB, Barbano HE, Cox CS, et al.Plan and operation of the NHANES I Epidemiologic Followup Study, 1982–84. *Vital Health Stat [22].* 1987; no. 1. DHHS publication no. PHS 87-1324.
10. Finucane FF, Freid VM, Madans JH, et al. Plan and operation of the NHANES I Epidemiologic Followup Study, 1986. *Vital Health Stat [25].* 1990; no. 1. DHHS publication no. PHS 90-1307.
11. Cox CS, Rothwell ST, Madans JH, et al. Plan and operation of the NHANES I Epidemiologic Followup Study, 1987. *Vital Health Stat [27].* 1992; no. 1. DHHS publication no. PHS 92-1303.
12. Pokras R, Hufnagel V. Hysterectomies in the United States, 1965–84. *Vital Health Stat [92].* 1987; no. 13. DHHS publication no. PHS 88-1753.
13. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley & Sons Inc; 1981.
14. *SAS Technical Report P-217 SAS/STAT Software: The PHREG Procedure. Version 6.* Cary, NC: SAS Institute Inc; 1991.
15. Horwitz RI, Yu EC. Problems and proposals for interview data in epidemiological research. *Int J Epidemiol.* 1985;14:463–467.

Recruitment Activities and Sociodemographic Factors That Predict Attendance at a Mammographic Screening Program

Susan F. Hurley, PhD, Richard M. Huggins, PhD, Damien J. Jolley, MSc, and Dorothy Reading, BA

ABSTRACT

A random sample of 2266 women aged 50 to 69 years was used to investigate factors that predict attendance at a free Australian mammographic screening program. The most important predictor was receipt of a personal invitation letter. A letter that included an appointment time increased attendance 132-fold initially and decreased to 20 times baseline after 14 days. A letter that did not include an appointment time increased attendance 12-fold, and a second letter to nonattenders increased attendance approximately 13-fold. Attendance declined with increasing distance from the program and with increases in the percentage of non-English speaking women in a neighborhood, but was higher in areas of higher socioeconomic status. (*Am J Public Health.* 1994;84:1655–1658)

Introduction

Women's adherence to recommended mammography screening schedules often is suboptimal,[1] and development of strategies that maximize participation is necessary for mammography to achieve its potential for breast cancer control.[2] We evaluated the impact of recruitment strategies and sociodemographic factors on attendance at the Essendon Breast X-ray Program, a pilot project conducted in the state of Victoria, Australia.

Methods

Target Population and Recruitment Strategies

During a 2-year period, the Essendon program offered free screening and consequential assessment to 43 771 women aged 50 to 69 years. These women lived in any of 34 surrounding postcode areas and were listed on the electoral roll,[3] for which enrollment is nominally compulsory.[4]

Susan F. Hurley was with the Department of Biostatistics, University of Washington, Seattle; she is now with the Department of Public Health and Community Medicine, The University of Melbourne, Carlton, Victoria, Australia. Richard M. Huggins is with the Department of Statistics, Latrobe University, Victoria, Australia. Damien J. Jolley is with the Cancer Epidemiology Centre, Anti-Cancer Council of Victoria, Australia. Dorothy Reading is with the Education Unit, Anti-Cancer Council of Victoria.

Request for reprints should be sent to Susan F. Hurley, PhD, Department of Public Health and Community Medicine, The University of Melbourne, 200 Berkeley St, Carlton, Victoria 3053 Australia.

This paper was accepted May 10, 1994.

TABLE 1—Attendance and Selected Sociodemographic Characteristics of the Sample, by Invitation Type

	No Letter (n = 832)	Letter A (n = 424)	Letter B (n = 1010)
Percentage who attended			
Before letter campaign	16.6	. . .	. . .
Total	43.8	40.1	26.6
Mean age in years (SD)			
Attenders	57.4 (5.9)	57.3 (5.8)	56.9 (5.6)
Nonattenders	59.5 (7.3)	58.2 (5.9)	57.7 (6.1)
Mean Indicator of Disadvantage[a] (SD)			
Attenders	99.1 (5.9)	96.5 (6.6)	98.6 (6.2)
Nonattenders	96.9 (5.8)	95.7 (6.2)	97.1 (6.7)
Attendance profile, %			
Attended after first letter	. . .	35.1	9.1
No further invitation	. . .	22.6	12.8
Follow-up letter and attended	. . .	5.0	17.5
Follow-up letter and did not attend	. . .	37.2	60.6

Note. Letter A included a specific appointment time, Letter B did not. The follow-up letter after Letter A or Letter B did not include a specific appointment time.

[a]The Indicator of Disadvantage was scored to have a median value of 100 for Australia. In the state of Victoria, the value ranged from approximately 80 to 120. For postcodes within the target area, the median was 96.2, mean was 97.7, and range was 84.5 to 113.7.

TABLE 2—Predictors of Attendance, from the Cox Proportional Hazards Model

Recruitment Activities and Sociodemographic Factors	Relative Risk[a]	95% Confidence Interval
Distance (km)[b]	0.97	0.966, 0.994
Percentage non-English speaking	0.99	0.988, 0.997
Indicator of Disadvantage[c]	1.03	1.02, 1.04
Letter A		
7–14 days	132	103, 170
15–42 days	19.6	12.5, 30.7
> 42 days	3.18	2.03, 4.99
Follow-up letter after Letter A		
1–49 days	1.56	0.70, 3.39
> 49 days	0.75	0.35, 1.63
Letter B		
1–42 days	12.2	9.2, 16.1
> 42 days	4.98	3.37, 7.36
Follow-up letter after Letter B		
1–49 days	13.25	3.62, 48.7
> 49 days	1.15	0.69, 4.96

[a]Each relative risk is adjusted for age and age squared as well as for the other variables in the table. Regression coefficients (standard errors) for age and age squared were 0.4056 (0.1295) and −0.364 (0.1120), respectively.

[b]The risk of attendance for a woman who lived 15 km from the program relative to a woman who lived 5 km from the program was $(0.97)^{10} = 0.74$ (i.e., a reduction of 26%).

[c]The risk of attendance for a woman living in a postcode with a value of 105 for the Indicator of Disadvantage (the 75th percentile for the target area) relative to a woman living in a postcode with a mean value of 93 (the 25th percentile for the target area) was $(1.03)^{12} = 1.43$ (i.e., 43% higher).

There was considerable diversity in socioeconomic status and ethnicity within the target area.

Both public and personal recruitment strategies were used to get the women to participate in the program. The public campaign included articles in local newspapers (delivered free to each home); community promotion (displays of multilingual information, establishment of appointment stalls in public places, and contact with community, workplace, and health information organizations by an education officer); and promotion of the program to primary care physicians by an education officer.

The personal campaign consisted of invitation letters sent to randomly selected women who had not attended. The women received either a letter stating a specific appointment time (Letter A) or a letter without a stated appointment time (Letter B). Random samples of women who had not attended or had not made an appointment 4 weeks after receiving the first letter were sent a follow-up letter without an appointment time. An information brochure and details of an interpreter service that was willing to translate the letter in five languages were included.

The three public strategies began at the start of the program, and at fairly low intensity, to give women who were motivated to attend the opportunity to do so. The invitation letter campaign began in the sixth month, and, thereafter, both personal and public recruitment strategies were used.

Analysis of Attendance

A random sample of 2266 women was selected. The sample was stratified by postcode of residence, 5-year age group, and invitation type (no letter, Letter A, and Letter B). We developed a multivariate survival model in which attendance for mammography was regarded as failure and observations of time until attendance were regarded as censored if a woman either turned 70 years of age, or died, or the program ended before the woman attended. The Cox proportional hazards regression model was used,[5] and analyses were performed with the P2L procedure in BMDP.[6] The proportional hazards model allows estimation of the risk of attendance for women with specific values of covariates (predictor variables) relative to the risk of attendance for women with different values of these variables. Covariates may be time-dependent (i.e., their value changes over time).

We investigated the following potential predictors of attendance: age; four postcode-level sociodemographic variables (distance from program, mean travel time by public transportation, the percentage of women who spoke a language other than English at home, and a census-based measure of socioeconomic status known as the Indicator of Disadvantage[7]); postcode-level variables summarizing the intensity of public recruitment activities

(number of local newspaper articles per month, hours of community promotion per month, and hours of promotion to physicians per month); and time-dependent variables based on the time in which invitation letters were sent.

Results

Attendance patterns are summarized in Table 1. On average, women who attended the program were slightly younger than nonattenders and lived in areas of higher socioeconomic status. Of women who were not sent an invitation letter, 43.8% attended, compared with 40.1% of women who were sent Letter A and 26.6% who were sent Letter B. Letter A plus a follow-up letter to nonattenders resulted in attendance of 42.7%, whereas attendance was 29.5% in the sample who received Letter B plus a follow-up letter to nonattenders. However, such simple comparisons of attendance rates are potentially misleading. Although the group that received no letter had the highest attendance rate, this does not mean that the public campaigns were effective, because 38% of the women from this group who attended did so when the public campaigns operated at low intensity and before the letter campaign began. These women were motivated to attend without having been exposed to recruitment campaigns. As the program progressed, women from the pool of women who had not attended were selected to receive an invitation letter.

The final Cox model is summarized in Table 2. Age and age squared were included because attendance rates increased with age until around the age of 56 years and then declined. The model suggests that a woman's risk of attending declined by approximately 3% in a multiplicative manner for every additional kilometer her postcode was located from the program and declined by 1% for every percent of the postcode's female population 50 to 69 years of age that spoke a language other than English at home. The relative risk for the Indicator of Disadvantage suggests that attendance was greater in postcodes with higher socioeconomic status, increasing by approximately 3% for every unit of the index.

After women received Letter A, attendance rates increased 132-fold in the first 7 to 14 days, then dropped to approximately 20 times baseline for 15 to 42 days and to approximately 3 times baseline after 42 days. The relative risk of attendance after receiving the follow-up letter after Letter A was 1.56 initially, dropping to 0.76 after 49 days, but the confidence limits indicate that attendance rates for women who received these letters did not differ significantly from baseline. Nevertheless, these variables were included in the model to provide a full description of the impact of personal invitations on attendance rates over time. After women received Letter B, attendance was 12 times baseline initially, dropping to approximately 5 times baseline after 42 days. A follow-up letter after Letter B increased attendance rates 13-fold initially (higher than the response to Letter B itself), but after 49 days attendance did not differ significantly from baseline. Variables summarizing the intensity of the three types of public recruitment were not statistically significant.

Discussion

Our findings regarding sociodemographic predictors are consistent with previous reports. A decrease in attendance with increasing distance from the screening site also was reported for a mobile service in the United Kingdom,[8] and the New York Health Insurance Plan trial found that women with shorter travel times were more likely to attend repeat screenings.[9] The lack of effect of public transportation travel time on attendance is not surprising, because 74% of women traveled to the program by car.[10] Cultural factors may have made women from non-English speaking backgrounds less motivated to attend, or the promotional material may have been less accessible, despite efforts to minimize this possibility. The lower attendance rate associated with living in a postcode with lower socioeconomic status is consistent with previous reports of lower participation for poorer and less educated women.[1,11]

The response to an invitation that included an appointment time was much higher than the response to a letter without a specific appointment time, a similar finding to a United Kingdom study.[12] However, an earlier analysis of data from our program showed that due to higher costs associated with Letter A, the most cost-effective recruitment strategy was a Letter B with a follow-up letter to nonattenders.[13]

Our analysis did not detect any effect of the public recruitment campaigns. In a previous analysis, Poisson regression analysis was used to investigate the role of public recruitment campaigns on postcode attendance rates early in the program for women who had not received a letter. No effect of promotion to physicians was detected, but local newspaper articles and community promotion did increase attendance rates.[13] Other studies have found that a doctor's recommendation increases participation in screening mammography,[1,14] and in the early phases of our program, 39% of women in a sample of 130 attended following their doctor's recommendation.[15] However, after education staff visits, doctors seemed only to encourage their patients to attend for short periods, perhaps explaining the overall lack of effectiveness of physician promotion. A similar lack of success in getting physicians to alter practice patterns in a community-wide intervention study was reported by Fletcher et al.[16]

The cause of the difference in results between our current and previous analyses regarding local newspaper articles and community promotion may be the fact that previous analysis considered only the first 16 months of the program or the fact that earlier analysis was based entirely on postcode-level predictors and attendance rates, whereas, in the current analysis, data on attendance and receipt of letters were available for individual women, but public recruitment information only was available at a postcode level. The large impact on attendance of a woman's receipt of an invitation letter is likely to have obscured any effect of promotion of the program within the woman's neighborhood.

Despite these limitations, our study has provided a useful description of factors that predict attendance in an Australian program. The most important predictor of attendance was receipt of an invitation letter, and attendance varied markedly with the type of letter. The Cox regression model may be useful for evaluation of other public health programs. □

Acknowledgments

This work was supported in part by the Victorian Health Promotion Foundation and the Anti-Cancer Council of Victoria. Susan F. Hurley is supported by a Neil Hamilton Fairley Fellowship from the National Health and Medical Research Council. Damien J. Jolley is supported by a Victorian Health Promotion Foundation Public Health Fellowship. We thank the staff of the Essendon Breast X-ray Program, particularly Ms Lichun Quang for programming assistance and the education staff for completing activity diaries.

References

1. The NCI Breast Cancer Screening Consortium. Screening mammography: a missed

clinical opportunity? Results of the NCI Breast Cancer Screening Consortium and National Health Interview Survey Studies. *JAMA*. 1990;264:54–58.
2. Hurley SF, Kaldor JM. The benefits and risks of mammographic screening for breast cancer. *Epidemiol. Rev.* 1992;14:101–130.
3. The Essendon Breast X-ray Program Collaborative Group. A mammographic screening pilot program in Victoria 1988–1990. *Med J Aust.* 1992;157:670–673.
4. Hurley SF. Screening: the need for a population register. *Med J Aust.* 1990;153: 310–311.
5. Cox DR. Regression models and life-tables (with discussion). *J R Stat Soc B.* 1972;34: 187–220.
6. Dixon WJ, Brown MB, Engelman L, et al. BMDP statistical software. Berkeley, Calif: University of California Press; 1985.
7. *Socioeconomic Indexes for Areas, Victoria 1986.* Canberra, Australia: Australian Bureau of Statistics; 1989.
8. Haiart DC, McKenzie L, Henderson J, et al. Mobile breast screening: factors affecting uptake, efforts to increase response and acceptability. *Public Health.* 1990;104:239–247.
9. Fink R, Shapiro S, Roester R. Impact of efforts to increase participation in repetitive screenings for early breast cancer detection. *Am J Public Health.* 1972;62:328–336.
10. Hurley SF, Livingston PM. Personal costs incurred by women attending a mammographic screening programme. *Med J Aust.* 1991;154:132–134.
11. Rutledge DN, Hartmann WH, Kinman PO, Winfield AC. Exploration of factors affecting mammography behaviors. *Prev Med.* 1988;17:412–422.
12. Williams EMI, Vessey MP. Randomised trial of two strategies offering women mobile screening for breast cancer. *BMJ.* 1989;299:158–159.
13. Hurley SF, Jolley DJ, Livingston PM, Reading D, Cockburn J, Flint-Richter D. Effectiveness, costs and cost-effectiveness of recruitment strategies for a mammographic screening program to detect breast cancer. *J Natl Cancer Inst.* 1992;84:855–863.
14. Lerman C, Rimer B, Trock B, Balshem A, Engstrom PF. Factors associated with repeat adherence to breast cancer screening. *Prev Med.* 1990;19:279–290.
15. Cockburn J, De Luise T, Hill D, Hurley S, Reading D, Russell I. Boosting recruitment to breast screening programmes [letter]. *Med J. Aust.* 1990;152:332.
16. Fletcher SW, Harris RP, Gonzalez JJ, et al. Increasing mammography utilization: a controlled study. *J Natl Cancer Inst.* 1993;85: 112–120.

The Accessibility of Drug Treatment for Pregnant Women: A Survey of Programs in Five Cities

Vicki Breitbart, MSW, Wendy Chavkin, MD, MPH, and Paul H. Wise, MD, MPH

ABSTRACT

Through simulated calls to 294 drug treatment programs in five cities, this study investigated access for pregnant women and compared New York City's provision of services in 1989 to that in 1993. In all sites, the majority of programs accepted pregnant women. There was also a marked improvement in the availability of services in New York City. Yet options were more limited for Medicaid recipients and women needing child care, and an appointment or referral for prenatal care was usually not offered. Although the door for treatment may be opening for pregnant women, institutional barriers still remain. (*Am J Public Health.* 1994;84:1658–1661)

Introduction

Recent reports estimate that as many as 8 million American women of childbearing age are using illicit drugs.[1] From 1986 to 1988 there was a documented 50% increase in the number of drug-exposed infants.[2] The responses to this crisis have been diverse. By 1992, 14 states had passed legislation promulgating substance abuse treatment or coordination of services for women.[3] However, 24 states have criminally prosecuted women for using drugs during pregnancy. The number of women prosecuted, who are mostly poor and African American, tripled between 1990 and 1992.[4]

Proponents of prosecution argue that addicted women must be compelled to seek treatment.[5,6] Yet in 1989, Chavkin conducted a simulated caller survey of drug treatment programs in New York City and found that more than half of the programs categorically excluded pregnant women and that the overwhelming majority refused treatment to crack-addicted women on Medicaid.[7] Targeting programs in five cities (including New York City), this present study similarly used simulated calls to assess whether exclusion based on pregnancy, method of payment, or the need for child care is widespread and whether it still prevails in New York City today. There are many complex factors that deter pregnant women from seeking substance abuse treatment; this survey focuses on major institutional barriers.

Methods

Site Selection

In addition to New York City, the sites for this study included Albuquerque, NM; Charleston, SC; Detroit, Mich; and Portland, Ore. They were selected on the basis of their geographic and demographic diversity, and their common concern with the issue of perinatal drug exposure. All the sites had conducted prevalence studies or had documented substance abuse by pregnant women.

Vicki Breitbart and Wendy Chavkin are with the Center for Population and Family Health at the Columbia School of Public Health, New York, NY. Paul H. Wise is with the Harvard Institute for Reproductive and Child Health and the Joint Program in Neonatology, Harvard Medical School, Boston, Mass.

Requests for reprints should be sent to Vicki Breitbart, MSW, Center for Population and Family Health, Columbia University, 60 Haven Ave B-3, New York, NY 10032.

This paper was accepted on March 16, 1994.

These sites also represent varied policy approaches toward the problem of drug use among pregnant women. Whereas Oregon has legislation that protects women against prosecution for drug use during pregnancy,[8] all the other sites are in states that have criminally charged women for using drugs during pregnancy.[4,8] In New York, after a dramatic upsurge in the number of "boarder babies" (drug-exposed infants in the hospital awaiting disposition by Child Protective Services), the State Department of Social Services altered the protocol regarding positive neonatal toxicology results. Drug exposure is still grounds for a report to Child Protective Services, but a full investigation must now assess parental fitness.[9]

Charleston has instituted a collaborative program between the obstetrical service in the university hospital and local law enforcement agencies, whereby any woman who presents for prenatal care and tests positive for drugs is reported to the local solicitor. Forty warrants have been issued to drug-using pregnant women, resulting in 25 arrests[10] (personal communication, Shirley Brown, Medical University of South Carolina, December 2, 1992). In contrast, university-based pediatricians, in Albuquerque, together with the local district attorney, established a program to provide treatment for any woman who is arrested and found to be addicted and pregnant.

Survey Questionnaire

A list of all the substance abuse treatment programs in each city was obtained from the state agency regulating drug treatment. Drug treatment experts in each site provided information about the local terminology and substance abuse pattern, which was incorporated into a script for the callers. Each caller described herself as 5 months pregnant, not receiving prenatal care, addicted, on Medicaid, and the mother of a 15-month-old child. She requested child care and recorded whether the program provided or arranged for the service. However, because many drug-using women do not seek prenatal care,[11] the caller did not try to solicit this service. Instead, she clearly stated that she had not received any prenatal care prior to the phone call and recorded every time she was either told that the program would provide such care on site or given a referral.

Callers identified themselves as seeking substance abuse treatment and talked to the receptionist or intake worker who provided admissions information. Two calls were made to every program, each by a different caller. If the information differed between the calls, a third call was made and the most frequent answer was selected. In all cases where the service was provided for free, it was added to the group that accepted Medicaid.

TABLE 1—Services Provided to Pregnant Women, by Drug Treatment Setting

Services Provided, by City	Outpatient			Residential/Detoxification		
	No. Sites	No. Accepted	% Accepted	No. Sites	No. Accepted	% Accepted
Accept pregnant women						
Albuquerque	20	20	100	1	1	100
Charleston	7	7	100	3	3	100
Detroit	68	41	60	8	8	100
New York City	117	104	89	37	22	59
Portland	25	24	96	8	6	75
Total	237	196	83	57	40	70
Accept pregnant women on Medicaid or for free						
Albuquerque	20	10	50	1	0	0
Charleston	7	5	71	3	2	67
Detroit	41	34	83	8	7	88
New York City	104	71	68	22	9	41
Portland	24	18	75	6	4	67
Total	196	138	70	40	22	55
Accept pregnant women and provide child care						
Albuquerque	16	3	19	1	0	0
Charleston	4	0	0	3	0	0
Detroit	38	7	18	8	5	63
New York City	104	26	25	22	0	0
Portland	20	3	15	6	3	50
Total	182	39	21	40	8	20

Results

In all, the survey included a total of 294 programs, although the number of programs in each city varied partly as a function of city size. Percentages presented here are based on the number of facilities that provided complete information in a category.

The majority of outpatient and residential programs—ranging from 64% of all programs in Detroit to 100% of all programs in Albuquerque and Charleston (Table 1)—accepted pregnant women. Method of payment, however, proved to limit access significantly. While all the outpatient programs in Albuquerque and Charleston accepted pregnant women, fewer of these programs accepted pregnant women on Medicaid (50% and 71%, respectively). This pattern prevailed in the outpatient programs in the other sites as well. Access to residential programs also decreased for women on Medicaid in all the sites; in Albuquerque, there was only one residential treatment program that accepted pregnant women, and it did not accept a Medicaid recipient.

If the woman was seeking child care, her options for drug treatment were severely restricted. Three quarters of the outpatient programs in New York City accepting pregnant women did not provide child care (Table 1). Child care was even less available at the other sites. Callers found no outpatient programs in Charleston and no residential or detoxification programs in Albuquerque, Charleston, or New York that responded to their child care needs.

More programs treating crack addiction were willing to accept pregnant women than were those treating heroin addiction—90% as compared with 73% (Table 2). Similarly, a higher proportion of programs accepting pregnant women and treating crack addiction accepted Medicaid (80%) than did those treating heroin (65%) or alcohol addiction (58%). More than a quarter of the crack programs accepting pregnant women provided child care, whereas the heroin and alcohol programs offered child care less than 20% of the time. Finally, although the callers clearly stated that they were pregnant and had not received prenatal care, in most instances the program

TABLE 2—Services Provided to Pregnant Women, by Drug

Services Provided, by City	Crack			Heroin			Alcohol			Other		
	No. Sites	No. Accepted	% Accepted	No. Sites	No. Accepted	% Accepted	No. Sites	No. Accepted	% Accepted	No. Sites	No. Accepted	% Accepted
Accept pregnant women												
Albuquerque	5	5	100	5	5	100	11	11	100	0	0	n/a
Charleston	8	8	100	1	1	100	0	0	n/a	1	1	100
Detroit	32	31	97	27	12	44	2	2	100	4	3	75
New York City	75	63	84	71	56	79	2	2	100	6	5	83
Portland	17	16	94	9	8	89	4	4	100	3	2	67
Total	137	123	90	113	82	73	19	19	100	14	11	79
Accept pregnant women on Medicaid or for free												
Albuquerque	5	3	60	5	2	40	11	5	45	0	0	n/a
Charleston	8	6	75	1	1	100	0	0	n/a	1	0	0
Detroit	31	27	87	12	10	83	2	2	100	4	3	75
New York City	63	50	79	56	34	61	2	2	100	5	4	80
Portland	16	12	75	8	6	75	4	2	50	2	2	100
Total	123	98	80	82	53	65	19	11	58	12	9	75
Accept pregnant women and provide child care												
Albuquerque	5	1	20	4	1	25	8	1	13	0	0	n/a
Charleston	6	0	0	0	0	n/a	0	0	n/a	0	0	n/a
Detroit	31	7	23	10	3	30	2	1	50	3	1	33
New York City	56	16	29	55	8	15	2	0	0	4	2	50
Portland	15	5	33	6	1	17	4	0	0	1	0	0
Total	113	29	26	75	13	17	16	2	13	8	3	38

TABLE 3—Comparison of Services at New York City Drug Treatment Programs in 1989 and 1993

Services Provided, by City	1989			1993		
	No. Sites	No. Accepted	% Accepted	No. Sites	No. Accepted	% Accepted
Accept pregnant women						
Methadone maintenance	31	21	68	55	47	85
Drug free	47	15	32	99	79	80
Total	78	36	46	154	126	82
Accept pregnant women on Medicaid or for free						
Methadone maintenance	21	21	100	47	47	100
Drug free	15	5	33	79	65	82
Total	36	26	72	126	112	89
Accept pregnant women and provide or refer for prenatal care						
Methadone maintenance	21	13	62	47	23	49
Drug free	15	2	13	79	40	51
Total	36	15	42	126	63	50
Accept pregnant women and provide or arrange child care						
Methadone maintenance	21	1	5	47	8	17
Drug free	15	1	7	79	23	29
Total	36	2	6	126	31	25

respondent did not provide an appointment or a referral for medical care.

Comparison of data from the two surveys of New York City programs indicated a dramatic shift between 1989 and 1993. Eighty-two percent of that city's programs accepted pregnant women in 1993 whereas only 46% had done so in 1989 (Table 3). There was also a marked change in every program parameter examined. More accepted Medicaid, more provided or referred for prenatal care, and more provided or arranged for child care, even though only one quarter of all facilities responded positively to this last request.

Discussion

In the five cities surveyed, the majority of substance abuse treatment programs accepted pregnant women. For New York City, this represents a significant increase in available services. We can only speculate as to whether this change resulted from legal pressure (there was a class action suit on behalf of women denied access to treatment because of pregnancy),[12] the availability of funds

from the state and federal government earmarked for this population, funding regulations and mandates, heightened awareness, or other factors. The data do demonstrate that New York City drug treatment programs have substantially altered their behavior toward pregnant women during these 4 years and are currently more likely to accept pregnant women. On the other hand, treatment options declined in all sites if the woman was a Medicaid recipient. Such differential access to treatment based on mode of reimbursement underscores the importance of including drug treatment services in the design of a national health care reform package.

Recent federal regulations for substance abuse treatment stress the importance of prenatal care and child care.[13] This study found little programmatic recognition of the client's need for these services. Prenatal care has been documented to have a beneficial impact on the birth outcomes of substance-abusing women.[14] Yet most staff who responded to our callers seemed unaware of the urgency of such a referral for a pregnant woman in her second trimester without medical care. More than a decade ago, researchers reported that lack of child care precluded addicted women's participation in treatment.[15] Unfortunately, this part of the picture has not changed; only a small proportion of programs make child care available.

Although the treatment door is now open to pregnant women, institutional barriers are still widespread, preventing poor women with children from getting care. Criminal prosecution and other similar policies that target the individual behavior of drug-using women implicitly discount the systemic inadequacies documented here. □

Acknowledgments

This research was funded by the Aaron Diamond, Ford, and Robert Wood Johnson Foundations. Special thanks to Dineen Gardner, Wendy Greene, Lisa Mauro-Bracken, Jeanette Rios, and Mark Rosing for their assistance with this project.

References

1. Adams EH, Gfroerer JC, Rouse BA. Epidemiology of substance abuse including alcohol and cigarette smoking. In: Hutchings DE, ed. *Prenatal Abuse of Licit and Illicit Drugs. Ann NY Acad Sci.* 1989;562:14–20.
2. *Drug Exposed Infants: A Generation at Risk.* Report to the Chairman of the US Senate Committee on Finance. Washington, DC: US General Accounting Office; June 1990.
3. *1992 Legislative Review.* Chicago, Ill: National Association of Perinatal Addiction Research and Education; 1992.
4. Paltrow L. *Criminal Prosecutions against Pregnant Women: National Update and Overview.* New York, NY: American Civil Liberties Union Foundation, Reproductive Freedom Project; April 1992.
5. *Commonwealth of Pennsylvania v Kemp.* Docket No. 114 (Pa Super Ct, Pittsburgh, 1993).
6. *Commonwealth of Kentucky v Welch.* No. 92-SC-490-DG (Ky Slip Op, September 30, 1993).
7. Chavkin W. Drug addiction and pregnancy: policy crossroads. *Am J Public Health.* 1990;80:483–487.
8. Government Relations Departments. *State Laws on Pregnant Women.* Washington, DC: American College of Obstetricians and Gynecologists; 1991.
9. Little R, Ensminger B. Implementing state policy regarding babies born with positive toxicologies, drug withdrawal syndromes, fetal alcohol effect, fetal alcohol syndrome or with any other condition which may be attributable to in utero exposure to drugs or alcohol. City of New York Human Resources Administration Memorandum; June 3, 1991.
10. Horger EO, Brown SB, Condon CM. Cocaine in pregnancy: confronting the problem. *J S C Med Assoc.* 1990;86(10):527–531.
11. Chavkin W, Paone D, Freidman P, Wilets I. Reframing the debate: toward effective treatment for inner city drug-abusing mothers. *Bull N Y Acad Med.* 1993;70(1):50–68.
12. *Elaine W. v North General Hospital.* 580 NYS 2d 246 (1992).
13. New requirements for substance abuse block grant recipients. Washington, DC: Legal Action Center Memo; January 29, 1992.
14. Allen M. Perinatal outcome after intense prenatal care for chemically dependent women. Annual meeting of the American College of Obstetricians and Gynecologists; May 1991; New Orleans, La. Abstract.
15. Reed B, Moise R. Implications for treatment and future research in addicted women. In: *Family Dynamics, Self-Perception, and Support System.* Washington, DC: National Institute on Drug Abuse; 1979; 114–130. DHEW publication ADM 80–762.

First Russian-American Breast Cancer Conference to Be Held

The first Russian-American Breast Cancer Conference will be held in Saratov, Russia, June 7–9, 1995. Organized to facilitate the sharing of information on advances in breast cancer diagnosis, treatment, and prevention, the conference is sponsored by the National Cancer Institute of the National Institutes of Health, US Department of State, National Alliance of Breast Cancer Organizations, University of North Carolina School of Medicine, Saratov Medical University, and Russia's National Oncology Center in Moscow. Nearly a dozen leading breast cancer experts from the United States will join specialists from throughout the former Soviet Union as conference leaders.

Although this conference was created primarily to update the Russian medical community, interest by North American and Western European physicians resulted in opening the conference to these physicians as well. This conference ends a half-century of seclusion for most doctors in the former Soviet Union from western scientific and medical development.

Conference arrangements are being coordinated by Svetlana Lisanti, President, Center for Bio-Medical Communication, Inc., Dumont, New Jersey. Additional conference information can be obtained by calling the center at (201) 385-8080.

National and State Spending on Specialty Alcoholism Treatment: 1979 and 1989

Joyce H. Huber, PhD, Gregory C. Pope, MS, and Debra A. Dayhoff, PhD

ABSTRACT

The National Drug and Alcohol Treatment Unit Survey was used to measure changes in specialty alcoholism treatment spending between 1979 and 1989 nationally and by state. National spending more than doubled from $1.6 billion to $3.8 billion in 1989 dollars. Private spending increased more rapidly than public spending, although most clients continue to be publicly funded. Dramatic differences across states in public funding growth were partially explained by differential increases in per capita income and in federal substance abuse block grants. Access to treatment continues to vary widely across the states. (*Am J Public Health.* 1994;84:1662–1666)

Introduction

During the 1980s, the financing of alcohol abuse treatment experienced a transformation in both the private and the public sectors. There was a dramatic increase in private third-party coverage of substance abuse treatment. In 1979, 17 states had legislation mandating that alcoholism treatment be offered by insurers as a benefit or an optional benefit.[1] By 1988, all but 10 states had enacted such legislation. Between 1981 and 1988, insurance coverage of alcohol abuse treatment increased from 36% to 83% of employees at medium and large private firms.[2]

Federal funding for substance abuse also underwent a transformation. Before 1982, the federal government funded substance abuse treatment through a series of categorical grants that targeted particular programs or populations. In 1982, these categorical grants were consolidated into the Alcohol, Drug Abuse, and Mental Health Services block grants, which are administered by the states. The initial 1982 block grant reduced funding by 25%.[3] Between 1985 and 1989, there was a real increase of 75% in federal block grant funds, including both the Alcohol, Drug Abuse, and Mental Health grants and the Alcohol and Drug Treatment and Rehabilitation grants.[4,5]

This paper compares funding and use of alcoholism treatment at specialized facilities in 1979 and 1989 to measure the changes resulting from the spread of private insurance coverage and consolidation of federal funding into the block grants.

Data and Methods

The primary data used for this study were the 1979 and 1989 National Drug and Alcoholism Treatment Unit Surveys, censuses of substance abuse treatment units sponsored by the National Institute on Alcohol Abuse and Alcoholism and the National Institute on Drug Abuse. In these surveys, all known facility-based units are requested to provide total annual funding for alcoholism and drug abuse treatment by source of funds, as well as capacity and number of clients treated on a particular date by modality (e.g., detoxification, rehabilitation) and setting (e.g., inpatient hospital, inpatient nonhospital, outpatient). Three types of treatment settings are not included in the surveys: treatment in nonspecialty beds of general hospitals, office-based treatment by private practitioners, and treatment through self-help groups such as Alcoholics Anonymous. These other settings, particularly self-help groups, may represent a substantial proportion of treatment for substance abuse or its problems.[6]

To improve estimates of total spending in 1979 and 1989, we imputed missing values for funding and used nonresponse weights to inflate our 1979 and 1989 funding, client, and capacity estimates. Missing values for funding were set to mean funding per unit capacity by ownership category (e.g., private for-profit, private nonprofit) and location (e.g., community mental health center, general hospital). We inflated funding, client, and capacity levels to compensate for survey nonresponse. In 1979, the overall response rate was 93.4%, but it dropped to 78% in 1989. There is no information for the 1979 survey on response rate by state or ownership type; therefore, we used the national nonresponse rate to inflate funding, capacity, and clients. We used state-specific nonresponse rates from 1989 and information on public and private nonresponse from a 1990 nonresponse survey[7] to adjust the 1989 levels.[8] All 1979 dollars were inflated to 1989 dollars by means of the Consumer Price Index. While nonresponse adjustments should produce accu-

The authors are with the Center for Health Economics Research, Waltham, Mass.

Requests for reprints should be sent to Joyce H. Huber, PhD, Center for Health Economics Research, 300 Fifth Ave, 6th Floor, Waltham, MA 02154.

This paper was accepted April 18, 1994.

Note. The views expressed are the authors' and do not necessarily represent the views of the National Institute on Alcohol Abuse and Alcoholism.

rate estimates of national growth in alcoholism treatment, they may result in biased estimates of growth in those states or ownership categories whose 1979 response rates differed substantially from the national response rate.

In the National Drug and Alcoholism Treatment Unit Survey, individual clients cannot be identified as publicly or privately funded. Clients in units with more than 75% of their funding from public sources were considered publicly funded clients, whereas clients in units with more than 75% of their funding from private sources were considered privately funded clients. If facilities received between 25% and 75% of their funding from public sources, clients were designated as publicly or privately funded in proportion to the funding from each source.

Results

National Trends in Alcoholism Funding and Clients at Specialty Treatment Units

The recognition of alcohol abuse as a serious, but treatable, medical and public health concern is reflected in changes in the levels and sources of funding between 1979 and 1989. Between 1979 and 1989, real spending on alcoholism treatment at specialty units more than doubled from $1.6 billion to $3.8 billion (Table 1). Private funding grew more rapidly than public funding, rising from one third to one half of the total. There was also a shift in both the private and public sectors to third-party funding. In 1979, public and private third-party insurance together accounted for about one quarter (28%) of total spending; by 1989 it accounted for nearly half (48%). Private third-party spending grew the most rapidly of any major category, as private insurance coverage for alcoholism treatment spread. Such spending accounted for $1 billion, or nearly half, of the $2.2 billion spending increase over the decade. Despite recent expansion of federal support for substance abuse funding through block grants, state-administered categories, including federal grants, have been the slowest growing of the major funding categories. However, the decreased importance of state and federal funds is partially offset by the increased importance of public third-party funding (primarily Medicaid). The shift away from alcoholism treatment as a welfare or philanthropic activity toward a medical insurance model of alcoholism treatment funding is reflected in both the increase in third-party funding and the erosion of the vestiges of funding from private donations and public welfare, which fell from 9% to 2.5% of total funding.

TABLE 1—Spending on Alcoholism Clients in Specialty Treatment Units, by Funding Source, 1979 and 1989

	1979[a]		1989		
	Thousands of Dollars	%	Thousands of Dollars	%	Change in Spending, %
Total spending	1 606 503	100.0	3 801 716	100.0	137
Total public spending	1 067 927	66.5	1 706 644	44.9	60
State[b] (including federal grants)	502 671	31.3	651 681	17.1	30
Local[c]	143 754	8.9	209 698	5.5	46
State/local[d] fee for service	24 117	1.5	155 900	4.1	546
Public welfare[e]	74 222	4.6	48 905	1.3	−34
Public third party[f]	137 331	8.5	445 041	11.7	224
Federal (direct)[g]	185 832	11.6	195 419	5.1	5
Total private spending	538 576	33.5	1 889 395	49.7	251
Private third party	316 603	19.7	1 379 389	36.3	336
Client fees	152 593	9.5	465 696	12.2	205
Private donations	69 380	4.3	44 310	1.2	−36
Other[h]			205 677	5.4	

Note. Adjustments were made for nonresponse and missing data.
[a]In 1989 dollars.
[b]Includes federal categorical grants in 1979 and federal block grants in 1989.
[c]City and county funds.
[d]State/local reimbursement to providers for services.
[e]Title XX, general assistance, food stamps.
[f]Champus (Civilian Health and Medical Program of the Uniformed Services), Medicare, Medicaid.
[g]Federal funds from agencies such as the Veterans Administration, the Drug Enforcement Administration, and the Bureau of Prisons.
[h]Not a category in 1979.

TABLE 2—Alcoholism Clients in Specialty Treatment Units, by Modality, 1979 and 1989

	1979		1989		Change in Clients, %
	No.	%	No.	%	
Total clients	306 772	100.0	511 240	100.0	67
Publicly funded clients	251 026	81.8	317 819	62.2	27
Privately funded clients	55 746	18.2	193 421	37.8	247
Inpatient hospital clients					
Total	20 589	6.7	26 262	5.1	28
Medical detoxification[a]	5 814	1.9	6 897	1.3	19
Residential clients					
Total	35 714	11.6	40 056	7.8	12
Rehabilitation/recovery	30 026	9.8	34 682	6.8	16
Outpatient clients					
Total	243 753	79.5	441 495	86.4	81
Rehabilitation/recovery	237 329	77.4	406 130	79.4	71
Other[b]	6 716	2.2	3 407	0.7	−49

Note. The table reflects the number of clients in treatment on April 30, 1979, and September 30, 1989. Adjustments were made for nonresponse and missing data.
[a]The main modalities were medical detoxification, social detoxification, and rehabilitation and recovery, sometimes referred to simply as alcoholism treatment.
[b]Other refers mainly to individuals whose location of treatment was a correctional facility. Clients in this category are not strictly comparable because of differences between the 1979 and 1989 survey forms.

TABLE 3—Public and Private Spending on Alcoholism Clients at Specialty Treatment Units, 1979 and 1989 (in 1989 Dollars)

State	Public Spending Per Capita, $ (Rank)			Private Spending Per Capita, $ (Rank)		
	1979	1989	Change, %	1979	1989	Change, %
United States	4.76	7.27	53	2.40	8.05	235
Alaska	24.17 (1)	35.73 (1)	48 (23)	1.52 (33)	15.74 (4)	939 (7)
District of Columbia	16.38 (2)	16.16 (4)	−1 (38)	1.66 (31)	9.60 (16)	478 (16)
Wyoming	16.02 (3)	12.56 (8)	−22 (49)	0.60 (46)	7.06 (28)	1079 (5)
Minnesota	10.40 (4)	10.19 (11)	−2 (40)	7.35 (3)	9.10 (19)	24 (47)
Montana	10.12 (5)	8.15 (16)	−19 (46)	2.48 (15)	9.22 (18)	271 (26)
New Mexico	9.82 (6)	13.56 (5)	38 (29)	0.48 (47)	5.93 (32)	1125 (4)
Colorado	9.18 (7)	7.22 (21)	−21 (48)	3.33 (9)	10.57 (14)	218 (31)
Oregon	8.45 (8)	5.24 (37)	−38 (51)	2.05 (25)	4.74 (37)	131 (41)
Wisconsin	8.08 (9)	9.49 (12)	17 (34)	5.34 (4)	11.39 (13)	113 (42)
Maine	7.66 (10)	10.78 (9)	41 (27)	2.28 (21)	18.79 (2)	726 (10)
Arizona	6.91 (11)	6.81 (27)	−1 (39)	4.09 (7)	3.42 (46)	−17 (51)
New York	6.88 (12)	18.86 (2)	174 (4)	2.22 (22)	13.52 (5)	508 (14)
Nebraska	6.80 (13)	6.89 (26)	1 (36)	3.25 (10)	10.03 (15)	209 (35)
South Dakota	6.74 (14)	7.16 (23)	6 (35)	1.73 (30)	5.12 (34)	196 (36)
Connecticut	6.62 (15)	13.21 (6)	100 (11)	2.44 (18)	7.74 (25)	218 (32)
Vermont	6.23 (16)	5.54 (33)	−11 (44)	0.62 (45)	4.72 (38)	657 (11)
Rhode Island	6.21 (17)	13.08 (7)	111 (8)	3.10 (13)	27.88 (1)	800 (8)
Massachusetts	5.92 (18)	9.19 (14)	55 (21)	2.67 (14)	9.07 (20)	239 (28)
North Carolina	5.80 (19)	5.68 (30)	−2 (41)	2.09 (24)	3.45 (45)	65 (46)
California	5.78 (20)	4.46 (45)	−23 (50)	4.10 (6)	7.88 (24)	92 (44)
Idaho	5.70 (21)	4.55 (44)	−20 (47)	0.70 (43)	11.77 (10)	1588 (2)
Kansas	5.29 (22)	6.94 (25)	31 (32)	3.15 (11)	13.16 (6)	318 (24)
Utah	5.03 (23)	7.43 (18)	48 (24)	2.04 (26)	9.49 (17)	366 (19)
Washington	4.91 (24)	9.29 (13)	89 (13)	3.11 (12)	11.88 (8)	282 (25)
Virginia	4.81 (25)	10.42 (10)	117 (7)	1.40 (35)	11.99 (7)	755 (9)
Arkansas	4.76 (26)	4.65 (42)	−2 (42)	0.81 (41)	6.15 (31)	655 (12)
South Carolina	4.62 (27)	7.27 (19)	57 (19)	0.28 (49)	3.55 (44)	1148 (3)
Florida	4.26 (28)	4.29 (46)	1 (37)	1.14 (38)	5.09 (35)	347 (22)
Iowa	4.08 (29)	16.97 (3)	316 (2)	2.48 (16)	6.22 (30)	151 (39)
Tennessee	4.03 (30)	3.73 (49)	−7 (43)	3.55 (8)	3.92 (42)	11 (49)
Pennsylvania	3.88 (31)	7.07 (24)	82 (15)	0.98 (40)	11.49 (11)	1070 (6)
Georgia	3.84 (32)	5.57 (32)	45 (26)	1.14 (37)	3.20 (47)	180 (37)
Illinois	3.79 (33)	5.52 (34)	46 (25)	2.40 (19)	7.44 (26)	209 (34)
North Dakota	3.56 (34)	7.23 (20)	103 (10)	10.04 (1)	11.86 (9)	18 (48)
New Hampshire	3.56 (35)	5.71 (29)	60 (18)	7.60 (2)	18.36 (3)	142 (40)
Michigan	3.52 (36)	5.27 (36)	49 (22)	4.34 (5)	7.90 (22)	82 (45)
Maryland	3.50 (37)	8.40 (15)	140 (5)	1.22 (36)	5.88 (33)	380 (18)
Louisiana	3.44 (38)	4.77 (39)	39 (28)	1.86 (28)	4.88 (36)	162 (38)
Kentucky	3.38 (39)	6.18 (28)	83 (14)	0.77 (42)	4.23 (40)	451 (17)
Nevada	3.25 (40)	3.82 (48)	18 (33)	1.96 (27)	4.00 (41)	105 (43)
Delaware	3.12 (41)	4.21 (47)	35 (30)	0.15 (51)	2.75 (48)	1689 (1)
Hawaii	3.07 (42)	5.14 (38)	67 (17)	2.16 (23)	2.11 (49)	−2 (50)
Texas	3.05 (43)	4.76 (40)	56 (20)	1.13 (39)	8.20 (21)	626 (13)
Indiana	2.77 (44)	5.31 (35)	91 (12)	1.83 (29)	6.72 (29)	268 (27)
Mississippi	2.66 (45)	4.69 (41)	77 (16)	0.27 (50)	1.16 (51)	324 (23)
Alabama	2.57 (46)	2.20 (51)	−14 (45)	1.47 (34)	4.62 (39)	215 (33)
Missouri	2.54 (47)	5.62 (31)	121 (6)	1.60 (32)	7.33 (27)	359 (21)
Oklahoma	2.46 (48)	7.96 (17)	224 (3)	0.48 (48)	1.59 (50)	232 (30)
West Virginia	2.34 (49)	3.08 (50)	32 (31)	0.63 (44)	3.73 (43)	492 (15)
Ohio	2.20 (50)	4.55 (43)	107 (9)	2.36 (20)	7.90 (23)	235 (29)
New Jersey	1.62 (51)	7.19 (22)	344 (1)	2.47 (17)	11.40 (12)	362 (20)
Coefficient of variation	70	67	138	83	61	95

Note. Adjustments were made for nonresponse and missing data.

State and local governments still predominantly fund treatment through programmatic grants but are relying more on fee for service reimbursement, perhaps to encourage greater accountability and productivity in treatment programs. Specialty alcoholism treatment remains heavily subsidized to individuals with only 9% and 12% of total funds coming from client fees in the 2 years.

The increase in real funding resulted in more access to treatment and greater intensity of treatment. The number of clients receiving treatment on a typical day rose from about 300 000 to about 500 000, a 67% increase (Table 2). Privately funded clients more than tripled, in comparison with a modest increase of 27% in publicly funded clients. However, on a typical day in 1989, the public sector still funded treatment for 50% more clients than the private sector. In 1979, 79.5% of clients were seen in outpatient settings; in 1989, 86.4% of clients were seen in such settings (Table 2). Correspondingly, the percentage of patients seen in both inpatient hospital and residential facilities has fallen.

Intensity of service, as measured by real spending per client-day, increased by 42% (from $14.35 in 1979 to $20.37 in 1989), in large part as a result of the relative shift of clients to the private sector, which provides more expensive treatment. In 1979, real spending per client-day was $11.66 for publicly funded clients, in comparison with $26.47 for privately funded clients. By 1989, spending per client-day had risen to $15.55 for publicly funded clients and $28.29 for privately funded clients, increasing 33% in the public sector and 7% in the private sector. Most of this differential spending growth can be attributed to a relatively greater growth of less costly outpatient treatment in the private sector. However, another part of the explanation may be a movement of individuals who are more treatable—those employed and with private health insurance—to the private sector, leaving only the more serious and costly cases in the public sector.

State Variations in Growth of Spending on Alcoholism Treatment at Specialty Units

Individual states have deviated considerably from national trends (Table 3). Despite significant growth nationally, 2 states experienced negative growth in private spending per capita, and 14 states experienced negative growth in public spending per capita. Six western states that were national leaders in 1979 saw public spending per capita fall by 19% to 36%: Oregon, California, Wyoming, Colorado, Idaho, and Montana. Two of these states, Idaho and Wyoming, had among

the highest growth rates in private spending per capita.

California experienced a 23% decrease in real public spending per capita and a 40% drop in publicly funded clients per capita (data not shown). California also had a below average increase in real private spending per capita. Consequently, the state fell from above to below the national average in alcoholism treatment funding per capita. The decrease in real public funding for alcoholism treatment can be largely attributed to a 66% drop in real local spending and a 41% drop in real public third-party (mostly Medicaid) spending. It seems likely that the California tax revolt, exemplified by Proposition 13, had a detrimental effect on alcoholism treatment spending.

Several states, on the other hand, made substantial gains in public financing of alcoholism treatment (e.g., New Jersey, Iowa, Oklahoma, New York, Maryland, and Rhode Island). New York evidenced a 174% increase in real public spending per capita, which was the fourth highest percentage increase in the country. Real public spending in New York rose from $121 million in 1979 to $338 million in 1989 and accounted for 34% of the national increase in public spending on alcoholism treatment (in comparison, the state comprises 7% of the nation's population). New York moved from 12th to 2nd in public spending per capita. Public third-party spending, primarily Medicaid, rose 307% from less than 3% of public spending on alcoholism treatment in 1979 to 46% in 1989. In a deliberate policy to use Medicaid to finance substance abuse treatment, New York expanded Medicaid coverage to individuals and services beyond federal eligibility requirements and foregoes federal matching funds for these individuals and services.[9]

Changes in state spending across the decade did not increase overall equity across states. Private spending per capita became more equitable, with the ratio of private spending per capita in the highest spending state to that in the lowest spending state falling by two thirds. However, variation in public spending per capita and total spending per capita (as measured by the coefficients of variation and the ratios of high to low) did not change significantly between 1979 and 1989. The federal block grant did not reduce the inequity across states in public funding per capita. The initial distribution of the block grant funds was based on historical funding levels. Although the allocation formula was revised several times to include states' measures of population and "need," these revisions applied only to incremental funding (most notably the Alcohol and Drug Treatment and Rehabilitation supplement established in 1987). Growth in nominal funding of the block grant was so low between 1979 and 1989 that the need-based incremental allocations had no distributional impact. Thus, there remained in 1989 considerable disparity across states in spending per capita, particularly for publicly funded clients.

TABLE 4—Regression of Changes in State/Local Government Funding for Alcoholism Treatment on Explanatory Variables

	Coefficient	*t*
Intercept	0.15	0.19
Change in per capita income[10,11]	0.54	3.11**
Change in per capita alcohol consumption[a,12]	1.26	1.16
Change in block grant 1985–1989[4,5]	0.80	1.79*
Mandated private insurance coverage of alcoholism treatment[1]	0.36	0.52
Adjusted R^2 = .20		
n = 50		

Note. All financial variables were measured in 1989 dollars. The dependent variable was change in state/local government spending on alcoholism treatment at specialty units, 1979–1989 (including federal grants to state and local governments), excluding Alaska.

Source. The original data on alcoholism treatment spending is from the National Drug and Alcoholism Treatment Unit Survey. The sources of original data for the remaining variables are indicated by the reference numbers.

[a]Other variables measuring change in need (i.e., rates of cirrhosis deaths, motor vehicle accidents, and alcohol-related arrests) were also included but had no effect on the growth of public spending on alcoholism treatment.

*$P < .10$; **$P < .01$.

Factors Influencing Growth of Public Spending on Alcoholism Treatment

Multivariate regression was used to examine factors that influenced growth in per capita state/local government spending on alcoholism treatment between 1979 and 1989 (Table 4). All state-administered spending, including federal substance abuse grants, was incorporated in the dependent variable. Enactment of mandated private insurance coverage of alcoholism treatment between 1979 and 1989 and change in alcoholism consumption per capita did not have a significant effect on public treatment spending. Conversely, each $1000 increase in real per capita income raised real per capita state/local government spending on alcoholism treatment by a statistically significant $0.54. Because higher federal grants may cause state governments to reduce their own appropriations, greater block grant funding does not necessarily raise total state-administered spending. However, the significant regression coefficient of .8 indicates that a $1.00 increase in the federal substance abuse block grant increased state-administered spending on alcoholism treatment by $0.80.

Conclusions

This paper has documented a dramatic increase in spending on specialty alcoholism treatment between 1979 and 1989, from $1.6 billion to $3.8 billion in 1989 dollars. The rate of spending growth has probably slowed in the early 1990s as many state and local governments have entered a period of fiscal crisis. In addition, third-party payers have reduced allowable reimbursements for substance abuse treatment, especially inpatient stays.

Funding for substance abuse treatment has been moving gradually toward an insured medical disease model. This trend may be brought to completion if substance abuse treatment benefits are included in the minimum benefits package for national health reform. Universal insurance coverage would be expected to raise overall spending on specialty substance treatment and the number of clients in treatment and to reduce inequities in access to care among the states. □

Acknowledgments

This paper was prepared for the National Institute on Alcohol Abuse and Alcoholism under grant 1R01AA08358-01.

We would like to thank Joy Camp of the Massachusetts Bureau of Substance Abuse Services for her clinical input and Helen T. Machado, Carol Ammering, and Sharron J. Bell for their production support.

References

1. Unpublished data. Chicago, Ill: Blue Cross/Blue Shield Association of America.
2. Morrisey M, Jensen G. Employer-sponsored insurance coverage for alcohol and drug abuse treatment, 1988. *Inquiry.* 1991; 28:393–402.
3. Institute of Medicine. *Broadening the Base of Treatment for Alcohol Problems.* Washington, DC: National Academy Press; 1990.
4. *State Resources and Services Related to Alcohol and Other Drug Abuse Problems, Fiscal Year 1985.* Washington, DC: National Association of State Alcohol and Drug Abuse Directors Inc; 1985.
5. *State Resources and Services Related to Alcohol and Other Drug Abuse Problems, Fiscal Year 1989.* Washington, DC: National Association of State Alcohol and Drug Abuse Directors Inc; 1989.
6. Schlesinger M, Dorwart RA, Clark R. Treatment capacity of drug problems in the U.S.: public policy in a fragmented system. In *Drug Policy and Health Services Research.* Bethesda, Md: National Institute on Drug Abuse; 1991.
7. *1990 National Drug and Alcoholism Treatment Unit Survey—Non-Response Analysis.* Bethesda, Md: National Institute on Drug Abuse/National Institute on Alcohol Abuse and Alcoholism; 1992.
8. Dayhoff DA, Pope GC, Huber JH. State variations in the public and private alcoholism treatment systems. *J Stud Alcohol.* 1994;55:549-560.
9. Larson MJ, Horgan CM. Variations in state Medicaid program expenditures for substance abuse units and facilities. In: *Financing Drug Treatment through State Programs.* Rockville, Md: National Institute on Drug Abuse; 1994. NIDA Treatment Services Research Monograph, no 1.
10. *Statistical Abstracts of the U.S.* Washington, DC: Bureau of the Census; 1982.
11. *Statistical Abstracts of theU.S.* Washington, DC: Bureau of the Census; 1991.
12. Alcohol Epidemiologic Data System. Washington, DC: CSR, Inc.

Monitoring Childhood Immunizations: A Canadian Approach

Janice D. Roberts, MB, MSc, Linda A. Poffenroth, MD, MSc, CommMed, Leslie L. Roos, PhD, Judith D. Bebchuk, MSc, and Anne O. Carter, MD

ABSTRACT

In Manitoba, Canada, a centralized, computerized childhood immunization monitoring system serves a population with insured medical coverage; each individual has a unique identification number. All physicians bill a single payer, and specific billing codes are used to identify immunizations given by physicians and by public health nurses. Together with dates of service, this information is used to construct immunizations-to-date and age-at-immunization profiles for individuals and groups. Reminders seeking missing information are built into the system and executed at strategic ages. The quality of input data has been assessed as high. The approach could be used readily by health maintenance organizations in the United States (*Am J Public Health.* 1994;84:1666–1668)

Introduction

The immunization program in Manitoba, Canada, is supported by a computerized information system (the Manitoba Immunization Monitoring System) that both measures and augments coverage among an entire population of children. The system is challenged by demography and delivery. While half the province's population of 1.1 million lives in Winnipeg, the remainder is scattered throughout rural and remote areas across 780 000 km^2 (300 000 square miles) and includes the residents of more than 60 Indian reserves. Except for Indians with treaty status, the provincial government is the single payer in a system of universal health insurance. Care for treaty-status Indians is the fiscal responsibility of the federal government. Three levels of government—municipal (city of Winnipeg), provincial, and federal—and two types of providers—physicians (80%) and public health nurses (20%)—are involved in the delivery of immunizations. Among children 14 years of age and under, we estimated that about 16% moved within Manitoba annually, while from 4% to 5% migrated into or out of the province.

Given the high level of American concern over immunization rates and their measurement, this paper offers a timely description of the Manitoba system and examples of the types of output available.

Methods

The Manitoba Immunization Monitoring System

Implemented in 1988, the Manitoba Immunization Monitoring System was constructed on the virtually complete population registry through which the province's health plan is administered; eligibility depends on registration. The

Janice D. Roberts, Linda A. Poffenroth, and Leslie L. Roos are with the Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada. Leslie L. Roos is also with the Manitoba Centre for Health Policy and Evaluation at the University of Manitoba. Judith D. Bebchuk is with the Outcomes Assessment Group, Duke University Medical Center, Durham, NC. Anne O. Carter is with the Department of Health Care and Promotions, Canadian Medical Association, Ottawa, Ontario, Canada.

Requests for reprints should be sent to Leslie L. Roos, PhD, Manitoba Centre for Health Policy and Evaluation, St. Boniface Hospital Research Centre, 351 Tache Ave, Rm R2008, Winnipeg, Manitoba R2H 2A6, Canada.

This paper was accepted July 11, 1994.

Note. The results and conclusions here are the authors', and no official endorsement by Manitoba Health is intended or should be inferred.

Editor's Note. See related editorial by Bernier (p 1567) in this issue.

computerized registry file includes, for each individual, a unique identification number, demographic identifiers, and dates of commencement and termination (migration out of the province or death) of plan enrollment. The file is regularly and automatically updated from the registration of adoptions, births, and deaths and from notification of migration into and out of Manitoba.

Children born on or after January 1, 1980, are automatically entered into the system at registration. The system file contains, for each child, registry data and the immunization record: four-digit codes identifying the (generic) vaccines administered and their sequence in the recommended immunization schedule, service dates, and the presence or absence of restrictions (to record vaccine contraindications).

Each child is assigned (by means of municipal and postal codes) to the public health office in the area of residence. Information concerning public health nurse immunizations is entered directly through terminals in all jurisdictions. Immunization data are transferred electronically to the system file from physician billing claims submitted to the health plan. Claims identify services (e.g., office visit, immunization) by tariff codes. Immunization codes represent vaccine type and sequence schedule (Table 1).

Immunization status is monitored by comparing, in the month of the first, second, fifth, and sixth birthdays, the system record and the recommended schedule. Missing or incorrectly coded immunizations produce a letter to the family (fifth birthday) or the provider (other birthdays) requesting correction or completion. "Reminders" are distributed through public health offices with amended records returned for data entry. Children whose records remain incomplete are actively followed by public health offices and offered immunization. Since the federal health department did not participate in the Manitoba Immunization Monitoring System until 1990, the records of Indian children have been considered incomplete and not yet monitored.

Comparisons between physician and system records have shown excellent agreement, with 2% or fewer immunizations coded incorrectly and service dates matching 98% of the time. The rate of physician failure to bill has been estimated at between 0.2% (urban) and 6.6% (rural). A direct examination of the efficacy of monitoring estimated that, among 1- and 2-year-olds, monitoring increased coverage by approximately 6%; missing immunizations were explained as failure to bill (43%), miscoding (27%), migration (20%), or no immunization (10%).

TABLE 1—Manitoba Immunization Monitoring System, Provincial Schedule of Routinely Recommended Immunizations, Acceptable Immunization Alternatives, and Tariff Codes

Tariff Code and Recommended Vaccine	Age	Tariff Code and Acceptable Alternative Vaccine
8601 DPT	2 mo	8641 DT or 8921 DPTP
8602 DTP	4 mo	8642 DT or 8922 DPTP
8603 DTP	6 mo	8643 DT or 8923 DPTP
8609 DTP	18 mo	8649 DT or 8929 DPTP
8509 DTP	5 y	8649 DT or 8929 DPTP
8659 Td	14 y	8609 DT or 8649 DT or 8929 DPTP
8611 OPV	2 mo	8921 DPTP or 8931 IPV
8612 OPV	4 mo	8922 DPTP or 8932 IPV
8613 OPV[a]	6 mo	8923 DPTP or 8933 IPV
8619 OPV	18 mo	8613 OPV or 8929 DPTP or 8939 IPV
8619 OPV	5 y	8929 DPTP or 8939 IPV
Hib[b]	2 mo	
Hib[b]	4 mo	
Hib[b]	6 mo	
Hib[b]	18 mo	
8670 MMR	12 mo	8621 measles
8661 rubella	12 mo	8670 MMR if given at or after 1 y of age

Note. DPT = diphtheria, pertussis, and tetanus vaccine; DT = diphtheria and tetanus toxoid; DPTP = diphtheria, pertussis, tetanus, and polio vaccine; Td = tetanus and diphtheria toxoid, "adult type"; OPV = oral polio vaccine; IPV = inactivated polio vaccine; Hib = *Hemophilus* b conjugate vaccine; MMR = measles, mumps, and rubella vaccine. The tariff code identifies the vaccine given by its generic type; for vaccines given in series, the last two digits identify the sequence. Physician billing claims use identical codes to represent immunization services.

[a]8613 OPV at 6 months of age is no longer recommended.

[b]Hib vaccine was not routinely recommended until 1992. Hib conjugate vaccine given in the same syringe with DTP: tariff codes 8781, 8782, 8783, 8789. All other Hib vaccines and Hib conjugate vaccine alone: tariff codes 8901, 8902, 8903, 8909.

TABLE 2—1989 Birth Cohort: Proportion with Immunization Schedule Complete, by Indian Status, Enrollment Status, and Place of Residence

	Infancy Schedule Recorded as Complete by First Birthday, % (No.)			
	Outside Winnipeg	Winnipeg		Total
		Inner City	Suburbs	
Non-Indian				
Continuous enrollment from birth to at least the first birthday	90.3 (6120)	87.3 (3273)	93.6 (5460)	90.8 (14 853)
Interrupted enrollment in the first year of life[a]	42.3 (442)	33.5 (340)	43.8 (377)	40.0 (1159)
Indian				
Continuous enrollment from birth to at least the first birthday[b]	38.6 (1519)	65.2 (368)	62.9 (70)	44.5 (1957)
Interrupted enrollment in the first year of life[ab]	12.8 (47)	30.0 (30)	50.0 (2)	20.3 (79)

[a]Recorded rates underestimate actual immunization rates because of migration.

[b]Recorded rates underestimate actual immunization rates because the federal health department did not fully participate in the Manitoba Immunization Monitoring System.

Population-Based Immunization Rates

System records of the total population of children born in 1989 and enrolled with the provincial Manitoba health plan in their first year of life (n = 18 048) were examined. Subgroups were defined by Indian status, enrollment status, and place of residence at the time of the initial Manitoba immunization. Recorded rates of completion of the recommended immunization schedule (three diphtheria–pertussis–tetanous immunizations and two oral polio vaccine immunizations, or appropriate alternatives [Table 1]) by the first birthday were calculated.

Results

For the total cohort, the recorded rate of schedule completion by the first birthday was 79%. Among children continuously enrolled during their first year of life, the completion rates were 91% for non-Indians (n = 14 853) and 45% for Indians (n = 1159) (Table 2). Recorded rates for children whose first-year enrollment was interrupted by migration (n = 1119) or by death (n = 119) were 40% and 20% for non-Indians and Indians, respectively.

The vast majority of children were continuously enrolled, the proportions being somewhat lower in Winnipeg's inner city (92%) than elsewhere (95%). The lowest completion rates recorded among continuously enrolled children were for non-Indians living in inner Winnipeg (87%) and for Indians living outside Winnipeg (39%).

Discussion

Like the Manitoba Immunization Monitoring System, computerized systems in use elsewhere measure the cumulative percentage of immunizations completed by certain ages in specified birth cohorts and calculate percentages of children up to date for recommended immunizations.[1,2] The Manitoba system, however, allows a closer look at the data; for example, calculation of the percentage of children receiving all scheduled immunizations within 30 days of the recommended age provides a measure of the proportion age-appropriately immunized.[3]

In addition to supplying measures of the quality of vaccination services, the system permits the definition of population subgroups by demographic characteristics and the calculation of rates by indicators (place of residence, Indian and enrollment status) of known influences on uptake: level of urbanization, access to service, and mobility.[4–7] Subgroups recording low coverage can be targeted and members tracked and offered immunization. Most children born in 1989 were up to date with recommendations by 1 year of age, regardless of area of residence. The low rates recorded among Indian children at least partially reflect jurisdiction-specific delays in data entry; manually recorded data from the federal health department suggest that coverage levels among Indian children are high. The effects of the rates recorded among Indians and migrants on overall rates are quantifiable and highlight the extent to which changes in the denominator influence reported coverage levels and limit comparisons between jurisdictions.[1,3]

The Manitoba Immunization Monitoring System is highly effective, relatively simple in design, and flexible to the needs of managers and users. It should be of interest to jurisdictions providing coverage of enrolled populations, such as American health maintenance organizations and other Canadian provinces. □

Acknowledgments

This work was supported in part by the Manitoba Centre for Health Policy and Evaluation through funding by Manitoba Health and by the Canadian Institute for Advanced Research. Dr Roos was funded by a National Health Scientist Award (NHRDP-6607-1314-48).

We extend thanks to Dr Margaret Fast, Dr Noralou Roos, Dr Barbara Law, and Dr David Fedson for their valuable comments. The assistance of Kerry Meagher and Randy Walld is acknowledged with thanks.

References

1. Begg NT, Gill ON, White JM. COVER (cover of vaccination evaluated rapidly): description of the England and Wales scheme. *Public Health.* 1989;103:81–89.
2. Verbrugge HP. The National Immunization Program of the Netherlands. *Pediatrics.* 1990;86(suppl):1060–1063.
3. Zell ER, Dietz V, Stevenson J, Cochi S, Bruce RH. Low vaccination levels of US preschool and school-age children. *JAMA.* 1994;271:833–839.
4. Baker MR, Bandaranayake R, Schweiger MS. Differences in rate of uptake of immunisation among ethnic groups. *BMJ.* 1984;288:1075–1078.
5. Bobo JK, Gale JL, Purushottam BT, Wassilak SGF. Risk factors for delayed immunization in a random sample of 1163 children from Oregon and Washington. *Pediatrics.* 1993;91:308–314.
6. Jarman B, Bosanquet N, Rice P, Dollimore N, Leese B. Uptake of immunisation in district health authorities in England. *BMJ.* 1988;296:1775–1778.
7. Jefferies S, McShane S, Oerton J, Victor CR, Beardow R. Low immunization uptake rates in an inner-city health district: fact or fiction? *J Public Health Med.* 1991;13:312–317.

Patient Adherence to Guidelines for Diabetes Eye Care: Results from the Diabetic Eye Disease Follow-Up Study

Julie C. Will, PhD, MPH, Robert R. German, MPH, Earl Schurman, Sharon Michael, RN, MS, Donna M. Kurth, and Larry Deeb, MD

ABSTRACT

Early detection and treatment of diabetic eye disease can prevent blindness, yet many persons with diabetes lack regular eye care. This study followed 569 people with diabetes participating in blindness prevention programs during 1985 through 1987; it was found that 35% of subjects received dilated eye examinations before entering the programs, in comparison with 60% afterward. About 85% of participants referred for proliferative retinopathy treatment began such treatment, and, of these, 85% completed treatment. A lack of knowledge about the disease and limited finances were primary reasons for nonadherence. To improve the effectiveness of prevention programs, eye care providers and program staff must strive to eliminate these educational and financial barriers. (*Am J Public Health.* 1994;84: 1669–1671)

Introduction

Diabetes is a leading cause of blindness in US adults.[1] Laser therapy applied to advanced diabetic proliferative retinopathy reduces the risk of severe vision loss by 50% or more.[2] Blindness prevention guidelines emphasize early detection of diabetic eye disease.[3–9] Recent studies indicate, however, that about a third of people with diabetes have never had ophthalmologic examinations and that more than half of these individuals have eye disease.[10,11] To improve the eye care of diabetic persons and prevent blindness, the Centers for Disease Control and Prevention (CDC) has worked with 27 state health departments to develop and implement blindness prevention programs.* Program staff have identified persons with diabetes who have not had a dilated eye examination within the previous year, have examined them to detect eye disease, and have educated them about the need for annual examinations and adherence to treatment recommendations. Here, we summarize the findings of the Diabetic Eye Disease Follow-up Study, which was designed to evaluate and improve the effectiveness of these state programs.

Methods

In 1988, the CDC requested proposals for participation in the Diabetic Eye Disease Follow-up Study from among the 10 state blindness prevention programs that had operated in the United States prior to 1987. Proposals were accepted from programs that had served at least 50 persons with diabetes before October 1, 1986, and were able to track these individuals. Four programs were eligible, and all were approved for funding by a CDC technical review committee. The study was approved by a CDC institutional review board.

Between 1988 and 1990, the study assessed the programs described in the Appendix. Between 1985 and 1987, 2106 people participated in the programs. Of these, we followed 569 individuals from a probability sample of Whites, Blacks, and Hispanics and a census of people from other racial/ethnic groups for 2 to 5 years. Nonparticipants were not identified or followed.

To select the sample, we employed stratified random sampling and used program records to define strata based on geographic location, urgency of treatment, and racial/ethnic group. Program participants in each location (i.e., Minnesota, Florida, Maryland, the eastern plains of Colorado, and inner-city Denver) were categorized into four groups by urgency of recommended treatment as defined by the clinical judgment of individual eye care specialists. Participants diagnosed with diabetic eye disease that required immediate treatment constituted the first group. Participants with diabetic eye disease that required reexamination in less than 1 year but not immediately composed the second group. Participants without diabetic eye disease who were

*CDC now works with 40 state and territorial health departments. Rather than provide eye care services, current programs concentrate on core public health functions (e.g., surveillance, quality assurance, access).

Julie C. Will and Robert R. German are with the Division of Diabetes Translation, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Ga. Earl Schurman is with the Local and Family Health Administration, Maryland Department of Health and Mental Hygiene, Baltimore. Sharon Michael is with the Division of Prevention Programs, Colorado Department of Health, Denver. Donna M. Kurth is with the Area Resource Center for Diabetes, Weiner Memorial Medical Center, Marshall, Minn. Larry Deeb is with the Department of Health and Rehabilitative Services, Tallahassee, Fla.

Requests for reprints should be sent to Julie C. Will, PhD, MPH, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Diabetes Translation, 4770 Buford Hwy, NE, Mailstop K-10, Atlanta, GA 30341-3724.

This paper was accepted February 15, 1994.

TABLE 1—Racial/Ethnic Characteristics of the US Diabetic Population, Program Participants, Individuals Selected for Study, Study Participants, and Subjects with Complete Medical Records

	White, %	Black, %	Hispanic, %	Other, %
US diabetic population (n = 4 778 000)	70.8	19.4	7.3	2.5
Program participants (n = 2106)	65.1	22.5	10.8	1.6[a]
Selected sample (n = 569)[b]	61.0	23.9	11.1	4.0
Completed interview (n = 414)	59.4	25.6	12.6	2.4
Complete medical record (n = 336)[c]	62.8	23.8	11.6	1.8

[a]Determined by use of program records. When interview data were used to determine racial/ethnic category, 10 of these individuals were later reassigned to the Black, White, or Hispanic category (the selected sample reflects this reassignment).
[b]One person was dropped from the sample because we were unable to definitively classify this individual by racial/ethnic category.
[c]The population of persons with diabetes who received a diagnostic examination and who had information recorded on the medical record abstraction form. These data were used to calculate treatment initiation and completion rates.

recommended for routine reexamination in 1 year constituted the third group. Finally, the fourth group included participants not classifiable by urgency of treatment. Individuals within each geographic area and group were then further stratified by racial/ethnic status. Within each area, all persons in the first group were selected for study (n = 172) to allow us to calculate treatment initiation and completion rates. The other groups were sampled to match the proportion of Blacks, Whites, and Hispanics in the first group.

Diabetic Eye Disease Follow-up Study investigators sent information to potential respondents, explaining the study's purpose and asking for permission to review medical records. Within 2 weeks, interviewers called subjects to request permission for a telephone interview. Interviewers made five attempts to contact each subject by telephone, made up to three visits to the subject's home, and, as a final attempt, sent the subject a questionnaire by mail. The 62-item, 20-minute questionnaire covered topics such as eye care history, knowledge, and barriers. Program eye evaluation records were then abstracted. If a medical release form was signed, the abstractor reviewed records from the subject's most recent eye care visit. The abstractor recorded the results of recommended treatment and reasons for not initiating treatment.

We employed various quality control procedures and used SUDAAN software to calculate proportions and standard errors and to conduct logistic regression analysis.[12] We used statistical weighting (corrected for nonresponse/lost to follow-up) to permit extrapolation to the total population of program participants.

Results

Four hundred forty-nine (79%) of the 569 subjects were contacted (complete medical records were obtained for 336 individuals, and interviews were completed for 414). One hundred twenty people (21%) were not contacted (74 had died, and 46 could not be reached).

Participants were more likely than the general diabetic population to be female, to use insulin, to be overweight, and to be financially disadvantaged (data not shown). They were also more likely to be Black or Hispanic (Table 1). The racial/ethnic status of persons who completed the interview or had complete medical records was not substantially different from that of all program participants.

We first determined the proportion of program participants initiating and completing treatment. As a result of program participation, 3% of persons were diagnosed with proliferative retinopathy, of whom 70% were referred for treatment. Treatment was initiated by 85% of the subjects referred, and, of these individuals, 85% completed the treatment. Of participants diagnosed with nonproliferative retinopathy, 5% were referred for treatment. Of the subjects referred, 48% initiated treatment, and of these individuals, 82% completed treatment.

Of the 74 individuals referred for treatment for proliferative or nonproliferative retinopathy, 25 (34%) did not initiate treatment within 1 year. Eye care records revealed that 12 (49%) of the 25 who did not initiate treatment were under the care of a doctor and were awaiting treatment; 6 (24%) persons indicated that they did not want treatment, and 1 (4%) had died. Reasons for not initiating treatment were not given in records for the remaining 6 subjects.

Second, we assessed the proportion of subjects who received annual examinations after joining the program. Sixty percent of interviewed participants reported receiving an annual examination during the year preceding the interview. Logistic regression analysis showed that proliferative and preproliferative retinopathy detected during the program's diagnostic examination strongly predicted later adherence to annual examination (odds ratio [OR] = 5.7, 95% confidence interval [CI] = 3.7, 8.8, for proliferative retinopathy; OR = 12.4, 95% CI = 7.0, 22.0, for preproliferative retinopathy). Also, the more recently patients participated in a blindness prevention program, the more likely they were to have annual examinations ($P = .002$).

Those who had not received an annual examination offered numerous explanations. The two most common reasons—given by approximately 60% of the respondents—were that they could not pay for an examination and that they had no visual symptoms.

Discussion

Evidence suggests that people who have never received eye care have difficulty adhering to eye care recommendations.[13] Only 35% of our subjects had received a dilated eye examination during the year prior to enrolling in a blindness prevention program. Despite this, 85% of program participants recommended for proliferative retinopathy treatment initiated and completed treatment. Furthermore, they continued receiving eye examinations. Two to 5 years after joining a program, 60% of participants reported having received an eye examination within the previous year. Indeed, diagnosis of retinopathy during program participation and recency of program participation strongly predict adherence to annual eye examination.

Given the effectiveness of retinopathy treatment, nonadherence to treatment recommendations signals a seriously

flawed system. Programs can improve this system by identifying people who do not initiate or complete retinopathy treatment and the reasons why they do not. Eye care providers should aggressively pursue and record the reasons for nonadherence and work with prevention program personnel to eliminate treatment barriers. Program personnel should consider establishing an annual examination reminder system to advise people with diabetes that retinopathy is often a silent disease and that annual examinations are important even when no visual symptoms are evident. Moreover, reminders should offer people information about low-cost services and other strategies to overcome financial barriers. □

Acknowledgments

We thank the interviewers—Donna Potter, Rosemary Hollick, Rahj Robinson, Peggy Seufert, and Sherry Tamminga—and the staff of the diabetes control programs in Florida (Kaye Harris, John Pattillo, Jamey Wise, Shary Gard, MS, and Jerry Shirah), Maryland (Susan Feldman), Colorado, and Minnesota (David Whipple). Finally, we acknowledge the support of Centers for Disease Control and Prevention staff members Mark Eberhardt, PhD, Frank Vinicor, MD, William Herman, MD, Frank DeStefano, MD, Scott Wetterhall, MD, MPH, David Olson, PhD, Braxton Moore, Ralph Yobp, Barbara Dougherty, Linda Geiss, MS, Alacia Lyons, RN, MA, and Myra Montalbano.

References

1. *Vision Problems in the U.S.* New York, NY: National Society to Prevent Blindness; 1980:1–46.
2. Diabetic Retinopathy Study Group. Photocoagulation treatment of proliferative diabetic retinopathy: clinical application of diabetic retinopathy study (DRS) findings. *Ophthalmology.* 1981;88:583–600.
3. *The Prevention and Treatment of Complications of Diabetes Mellitus: A Guide for Primary Care Practitioners.* Atlanta, Ga: Centers for Disease Control; 1989.
4. *The Physician's Guide to Type I Diabetes (IDDM).* Alexandria, Va: American Diabetes Association; 1988.
5. *The Physician's Guide to Type II Diabetes (NIDDM): Diagnosis and Treatment.* Alexandria, Va: American Diabetes Association; 1988.
6. American Diabetes Association. Guidelines for eye care in patients with diabetes mellitus. *Diabetes Care.* 1988;11:745–746.
7. *Diabetic Retinopathy Guidelines.* Lansing, Mich: Michigan Department of Public Health; 1984.
8. *Preferred Practice Patterns.* San Francisco, Calif: American Academy of Ophthalmology; 1989.
9. *Information on the Care of Diabetic Patients.* Alexandria, Va: American Optometric Association Board of Trustees; 1987.
10. Sprafka JM, Fritsche TL, Baker R, Kurth D, Whipple D. Prevalence of undiagnosed eye disease in high-risk diabetic individuals. *Arch Intern Med.* 1990;150:857–861.
11. Witkin SR, Klein R. Ophthalmologic care for persons with diabetes. *JAMA.* 1984;251: 2534–2537.
12. Shah BV. *SUDAAN: Program for Computing Standard Errors of Standardized Rates from Sample Survey Data.* Research Triangle Park, NC: Research Triangle Institute; 1991.
13. Newcomb PA, Klein R. Factors associated with compliance following diabetic eye screening. *J Diabetes Complications.* 1990;4: 8–14.

APPENDIX—Descriptions of the Four Blindness Prevention Programs Used in the Study

Maryland

The Maryland program employed a two-person team from the Johns Hopkins Diabetes Center to conduct eye disease screenings at 14 sites throughout a wide geographic area. Free screenings were conducted at community health centers, hospitals, local health departments, and local American Diabetes Association chapters. Those who had not received a dilated eye examination during the previous year were asked to identify themselves by responding to newspaper and public service announcements and to promotions at local diabetes education classes and American Diabetes Association chapter meetings. Identified persons were then recruited for nonmydriatic fundus photography of each eye, visual acuity testing, tonometry, blood pressure measurement, and eye care education. Seven hundred seventy-three persons with diabetes participated between 1985 and 1987. The photographs were read by a retinal specialist at Johns Hopkins University. Patients were referred to an ophthalmologist if proliferative retinopathy was detected. All other patients were urged to seek diagnostic examinations.

Florida

People with diabetes were examined for eye disease at two Pinellas County public health units located in St. Petersburg and Clearwater. This program was available to public health clinic clients with diabetes who had not received a diagnostic eye examination within 12 months prior to the screening examination. Local print, radio, and television media participated in announcing the availability of screening services for public health clients. Five hundred seven persons with diabetes participated between 1985 and 1987. The examination included visual acuity testing, intraocular pressure measurements, nonmydriatic retinal photographs of both eyes, and eye care education. Retinal photographs were sent to a university-based school of ophthalmology for evaluation. Depending on ophthalmologists' judgments of the severity of disease, patients were referred either for a confirmatory diagnostic examination or for routine follow-up within 1 year.

Minnesota

An epidemiologic study in Marshall enumerated all individuals with physician-defined diabetes who used the local medical care system (n = 533). Those who reported never having had an ophthalmologic examination (n = 172) were referred for a free eye examination at a local ophthalmologist's office. One hundred forty-five persons (85%) received the examination, which included visual acuity testing, tonometry, dilated funduscopic examinations, and mydriatic fundus photography. If proliferative retinopathy with high-risk characteristics was detected by the ophthalmologist, patients were given immediate treatment. Those with less serious eye disease were recommended for follow-up. Patients were also given eye care education at a community-based education program located at the same health care center as the ophthalmologist's office.

Colorado

The Colorado program targeted residents of inner-city Denver and the rural eastern plains. The inner-city residents generally have low incomes and traditionally do not use preventive health practices for diabetes control. Diabetes status was determined by reviewing medical and billing records of a major inner-city medical facility. The medical charts of all persons with diabetes were then identified by affixing brightly colored stickers to the patient's problem list. Health care providers were asked to use these charts to identify persons with diabetes and to refer those who had not had an eye examination during the past year to the medical facility's eye clinic. Residents of the eastern plains are dispersed throughout a large geographic area that has limited ophthalmologic services. These program participants were recruited through public announcements and by asking primary care physicians and health facilities to refer persons with diabetes to the program. Program publicity included announcements via radio and newspapers, hospitals, and pharmacies. Announcements emphasized that persons with diabetes who had not had a dilated eye examination during the past year were eligible for the program. All recruited persons were examined by an ophthalmologist who was in private practice or was temporarily located in a special clinic. Six hundred eighty-five persons with diabetes participated in the program between 1985 and 1987.

The Risk of Hospitalization for Ischemic Heart Disease among Asian Americans in Northern California

Arthur L. Klatsky, MD, Irene Tekawa, MA, Mary Anne Armstrong, MA, and Stephen Sidney, MD

ABSTRACT

To supplement sparse data about ischemic heart disease in Asian Americans, hospitalization risk was prospectively examined in a group of Asian Americans living in Northern California. Analyses used Cox models with ischemic heart disease risk traits as covariables. With Whites as the referent category, relative risks by ethnic group were as follows: Chinese, 0.6 ($P = .001$); Japanese, 1.0 ($P = .97$); Filipinos, 1.0 ($P = .95$); South Asians, 3.7 ($P < .001$); other Asian Americans, 0.8 ($P = .55$). Thus, unexplained differences in ischemic heart disease risk exist among Asian-American ethnic groups, with Chinese Americans at lowest risk and persons of South Asian origin at highest risk. (*Am J Public Health.* 1994;84:1672–1675)

Introduction

Considering the recent rapid increase in numbers of Asian Americans, data about health risks in this group remain scanty.[1–3] Prospective data about ischemic heart disease incidence are limited to epidemiological surveys among Japanese Americans residing in the United States.[4–6] These surveys show more ischemic heart disease among Japanese Americans than among Japanese living in Japan, a finding attributed to life-style changes. Studies of ischemic heart disease risk factors among Chinese Americans suggest that their risk is probably low.[7,8] Two reports suggest a high prevalence of hypertension among Filipino Americans.[9,10] Reports of a high ischemic heart disease prevalence among overseas Asian Indians in various countries have appeared.[11–18] To address the clear need for more data about ischemic heart disease in Asian-American ethnic groups, we here report prospective data about hospitalization for this condition.

Methods

We used data from 13 245 Asian Americans who voluntarily took health examinations offered by a Northern California prepaid health care program from 1978 through 1985. The self-administered questionnaire included the question "What is your race?" (Options: Black, White, Oriental [Asian], Latin, Mixed, and Other.) Subheadings under Oriental (Asian) were Chinese (n = 6064), Japanese (n = 1718), Filipino (n = 4283), and Other Asian (n = 1180). A separate query about birthplace yielded 274 South Asians ("Other Asians" born in India or Pakistan); the remainder are called "Other Asians" in this paper. The study population and examination procedure, which included health measurements and blood tests, have been described elsewhere.[8,19] Hospitalizations for ischemic heart disease were identified primarily by hospital discharge codes 410 through 414 (*International Classification of Diseases*, Ninth Revision[20]) in records of hospitals used by the health care program.

The subjects were followed from the date of first examination until December 1989, termination of health plan membership, or first hospitalization for ischemic heart disease (whichever was earliest). This procedure yielded 207 persons with such hospitalizations (99 for acute myocardial infarction and 108 for other ischemic heart disease diagnoses). Cox proportional hazards models were used for analyses, with age, sex, smoking (never smoker, ex-smoker, <1 pack/day, ≥1 pack/day), body mass index, marital status (married, never married, formerly married), education (no college, some college, college graduate), and alcohol intake (never drinker, ex-drinker, <1 drink/day, ≥1 drink/day) as covariates. Systolic blood pressure, total serum cholesterol, and random blood glucose were also included in some models. The use of these two models enabled study of the role of these established ischemic heart disease risk factors in the results. Comparisons were made between each Asian-American group and Whites (using a 10% random sample of Whites as the referent; n = 7206, of which 212 had been hospitalized for ischemic heart disease) and between the Asian-American groups themselves (using the largest subset, Chinese, as the referent). The results are expressed as relative risks (RRs) for hospitalization for ischemic heart disease.

Of the 13 245 Asian-American subjects, 33.1% terminated membership in the health care program by December 1989. By ethnic group, the percentages were Chinese, 33.8%; Japanese, 33.8%; Filipinos, 28.3%; South Asians, 44.5%; other Asians, 46.4%; Whites, 35.9%.

The authors are with the Division of Research, Kaiser Permanente Medical Care Program, Oakland, Calif; Dr Klatsky is also with the Division of Cardiology.

Requests for reprints should be sent to Arthur L. Klatsky, MD, Division of Cardiology, Kaiser Permanente Medical Center, 280 W MacArthur Blvd, Oakland, CA 94611.

This paper was accepted March 16, 1994.

TABLE 1—Traits of Subjects in Asian-American Ethnic Groups and Whites

	Chinese	Japanese	Filipinos	South Asians	Other Asians	Whites (10% Sample)
No. (% of total)	6064 (46)	1718 (13)	4283 (32)	274 (2)	906 (7)	7206 (. . .)
Mean age, y (SD)	39.1 (13.9)	41.5 (13.7)	37.7 (12.4)	35.6 (9.5)	34.5 (10.5)	42.5 (15.6)
Male, %	46.3	40.6	41.0	58.8	42.5	47.4
US-born, %	42.5	78.5	8.3	0[a]	11.3	88.3
Current smokers, %	10.9	20.1	18.7	13.9	23.1	23.1
College graduates, %	46.8	51.3	50.3	65.0	40.9	44.2
Married, %	61.8	60.7	68.3	79.9	66.0	49.0
Lifelong alcohol abstainers, %	34.1	15.9	35.2	29.2	31.6	6.2
Mean cholesterol, mmol/L	5.45	5.74	5.48	5.34	5.25	5.57
Mean glucose, mmol/L	5.3	5.4	5.4	5.4	5.2	5.3
Mean body mass index[b]	22.2	22.8	23.3	23.4	22.8	24.3
Mean systolic blood pressure, mm Hg	120.2	120.6	122.9	119.6	117.0	125.3
Ischemic heart disease risk "yes,"[c] %	26.5	30.4	40.2	29.9	31.1	32.8
Mean follow-up, y	7.3	7.4	7.4	6.4	6.2	7.1

[a]All persons in this subset stated that they were Asian and were born in India or Pakistan.
[b]Weight in kilograms divided by height in meters squared.
[c]See Methods for definition.

We attempted to identify persons who were free of or at low risk for ischemic heart disease. A person was considered at risk if he or she answered "yes" to any of 12 health history questions suggesting lifetime history of ischemic heart disease, symptoms of ischemic heart disease, or risk factors for ischemic heart disease (hypertension, diabetes, hypercholesterolemia).

Results

Tables 1 and 2 present descriptive data and age-adjusted hospitalization rates for the Asian-American groups. Age-specific rates for Chinese men were roughly half those of Whites; rates for South Asian men were two to four times those of Whites. For example, rates per 1000 person-years among men aged 45 through 59 years were 5.6 for Whites, 2.9 for Chinese, and 21.9 for South Asians; among men aged 60 years or older, rates were 16.0 for Whites, 8.3 for Chinese, and 34.7 for South Asians.

The risk of hospitalization, adjusted for risk factors, was similarly lower in Chinese Americans and higher in South Asians than in Whites (Table 3). Separate analyses by sex showed the following statistically significant relative risks (with 95% confidence intervals [CIs]): Chinese vs White men, 0.6 (0.4, 0.8); South Asian vs White men, 4.4 (2.1, 8.7); Filipino vs White women, 2.3 (1.2, 4.2). Other comparisons by sex of Asian subsets vs Whites were not significant; there were no cases of ischemic heart disease among South Asian women and therefore no risk ratios. The covariates showed expected relationships to ischemic heart disease risk. In the model with Whites as the referent, the following relative risks were found: age, 1.8 per 10 years ($P < .001$); male sex, 3.5

TABLE 2—Age-Adjusted Rates of Hospitalization[a] for Ischemic Heart Disease in Asian Americans and Whites, by Sex

	Men			Women		
Ethnicity	n	No. Hospitalized	Rate	n	No. Hospitalized	Rate
Chinese	2805	55	1.6	3259	17	0.4
Japanese	697	28	2.8	1021	9	0.6
Filipino	1756	55	3.4	2527	27	1.3
South Asian	161	9	9.0	113	0	. . .
Other Asian	385	7	2.9	521	0	. . .
White	3414	145	3.0	3792	67	1.0

[a]Age-adjusted rate per 1000 person-years, using age distribution of all Asians as standard population.

TABLE 3—Adjusted Relative Risk of Hospitalization[a] for Ischemic Heart Disease: Asian Americans vs Whites

Ethnicity	Relative Risk	95% Confidence Interval	*P*
White (referent)	1.0	. . .	. . .
Chinese	0.6	0.4, 0.8	.001
Japanese	1.0	0.7, 1.4	.97
Filipino	1.0	0.7, 1.4	.95
South Asian	3.7	1.8, 7.4	< .001
Other Asian	0.8	0.3, 1.8	.55

[a]Computed by Cox proportional hazards models controlled for age, sex, smoking, body mass index, marital status, education, alcohol intake, systolic blood pressure, total serum cholesterol, and blood glucose.

TABLE 4—Adjusted Relative Risk of Hospitalization[a] for Ischemic Heart Disease: Asian-American Ethnic Groups vs Chinese Americans

Ethnicity	Relative Risk	95% Confidence Interval	P
Chinese (referent)	1.0	...	...
Japanese	1.8	1.2, 2.7	.006
Filipino	1.9	1.3, 2.7	<.001
South Asian	6.6	3.2, 13.6	<.001
Other Asian	1.3	0.6, 3.1	.49

[a]Computed by Cox proportional hazards models controlled for age, sex, smoking, body mass index, marital status, education, and alcohol intake.

($P < .001$); body mass index, 1.1 per unit (kg/m^2) ($P < .001$); current smoking of 1 or more packs per day (vs never smoking), 1.8 ($P = .002$); current alcohol intake of 1 or more drinks per day (vs never drinking), 0.4 ($P < .001$); systolic blood pressure, 1.1 per 10 mm Hg ($P < .001$); total cholesterol, 1.1 per 0.25 mmol/L ($P < .001$); glucose, 1.03 per 0.5 mmol/L ($P < .001$). Compared with the Chinese, the Japanese, Filipino, and South Asian groups were all at higher risk, with a striking relative risk for South Asians vs Chinese (Table 4). The confidence intervals were wide for the comparisons of South Asians vs Whites and South Asians vs Chinese because of small numbers, but the P values for both comparisons were <.001. These results were little affected by inclusion of cholesterol, blood pressure, and glucose in the analyses. Addition of the ischemic heart disease risk variable (for "yes" vs not "yes," RR = 2.6, 95% CI = 2.1, 3.4) to the multivariate model resulted in only minor changes in relative risk estimates: for example, for Chinese vs Whites, RR = 0.6 either way; for South Asians vs Whites, RR = 3.7 with no control for ischemic heart disease risk and 3.8 with control for ischemic heart disease risk.

Discussion

To our knowledge, this is the first report of differences in ischemic heart disease risk that includes data for most of the major Asian-American ethnic groups living in the United States. Hospitalization represents only part of the clinical ischemic heart disease spectrum. This could be an important limitation if there are ethnic differences in the decision to hospitalize a patient who has experienced an ischemic heart disease event. We believe this is unlikely because of probable, similarly high identification of ischemic heart disease events among ethnic groups in this study sample. All subjects were sufficiently health conscious to take a health examination. In this study the ischemic heart disease risk of Japanese and Filipino Americans was similar to that of Whites. The data for Japanese Americans are consistent with extrapolation of the apparent effects of acculturation on ischemic heart disease risk in this group seen in previous studies.[4–6] Nearly 80% of Japanese Americans in this study population were US-born, and their traits were similar to those of US Whites.[8,19] Although 90% of the Filipino Americans in this study were born in the Philippines, the high prevalence of hypertension[9,10] and relatively high levels of ischemic heart disease risk factors in this group[8] seem sufficient to explain their ischemic heart disease risk. Efforts toward ischemic heart disease risk factor control in these groups should probably parallel those among US Whites and Blacks.

A lower ischemic heart disease risk in overseas Chinese has been reported in a multiethnic Singaporean population, without explanation.[13] A prediction of lower risk in Chinese Americans in this study population[8] was based on its lower prevalence of established ischemic heart disease risk factors. Our finding of lower risk for the Chinese Americans in the present study, independent of the controlled risk factors, does not rule out this explanation. Measurement of traits at a single point in time does not control for lifelong risk. Uncontrolled factors (e.g., high-density lipoprotein [HDL] cholesterol, physical activity, psychological traits) may be involved. We hope that others will explore these aspects. With further follow-up, larger numbers of subjects with ischemic heart disease will enable separate study of Chinese Americans born in China, Taiwan, Hong Kong, Hawaii, and the mainland United States and may provide further clues about ischemic heart disease risk in this ethnic group.

Our data show that Asian Americans of South Asian origin are at very high risk for ischemic heart disease. Although the number of cases is small (with resultant wide confidence intervals), the validity of the finding is strongly supported by reports from other countries.[10–18] Consistency in diverse environments suggests that an intrinsic genetic or cultural trait is involved. An explanation in terms of established risk factors is not yet clear. There seem to be large disparities in incidence of ischemic heart disease in various portions of India.[18] Others have considered the most promising explanatory hypothesis to be high prevalence of diabetes mellitus, central obesity, and insulin resistance.[12,15–18] Our data, controlled for blood glucose and body mass index, do not support this hypothesis, but we unfortunately have no data about HDL cholesterol, apolipoproteins, lipoprotein(a), or other possible lipoprotein factors. Our control for body mass index is also an inadequate test of this hypothesis, in the absence of data about skinfold thickness or waist-to-hip ratios. It is apparent that further research about the epidemic of ischemic heart disease in overseas South Asians is an urgent public health need. Whatever the explanation may prove to be, aggressive evaluation and modification of risk for ischemic heart disease in South Asian Americans seems appropriate. □

Acknowledgments

We thank Cynthia Landy and Rita J. Coston for data collection, Harald Kipp for assistance in preparing the data set, and Yolanda Yarbough for technical assistance.

References

1. Yu YS. The health risks of Asian Americans. *Am J Public Health.* 1991;81:1391–1393.
2. Lin-Fu JS. Population characteristics and health care needs of Asian Pacific Americans. *Public Health Rep.* 1988;103:18–28.
3. Yu ESH, Liu WH. US National Health Data on Asian Americans and Pacific Islanders: a research agenda for the 1990s. *Am J Public Health.* 1992;82:1645–1652.
4. Gordon T. Mortality experience among the Japanese in the United States, Hawaii, and Japan. *Public Health Rep.* 1957;72:543–553.
5. Robertson TL, Kato H, Rhoads CG, et al. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California. Incidence of myocardial infarction and death from coronary artery heart disease. *Am J Cardiol.* 1977;39:239–243.
6. Yano K, MacLean CJ, Reed DM, et al. A

comparison of the 12-year mortality and predictive factors of coronary heart disease among Japanese men in Japan and Hawaii. *Am J Epidemiol.* 1988;127:476–487.
7. Choi ESK, McGandy RB, Dallal GE, et al. The prevalence of cardiovascular risk factors among elderly Chinese Americans. *Arch Intern Med.* 1990;150:413–418.
8. Klatsky AL, Armstrong MA. Cardiovascular risk factors among Asian Americans living in Northern California. *Am J Public Health.* 1991;81:1423–1428.
9. Angel A, Armstrong MA, Klatsky AL. Blood pressure among Asian-Americans living in northern California. *Am J Cardiol.* 1989;64:237–240.
10. Stavig GR, Igra A, Leonard AR. Hypertension among Asians and Pacific Islanders in California. *Am J Epidemiol.* 1984;119:677–691.
11. Fox KM, Shapiro LM. Heart disease in Asians in Britain. *BMJ.* 1988;297:311–312.
12. Miller GJ, Beckles GL, Maude GH, et al. Ethnicity and other characteristics predictive of coronary heart disease in a developing community: principal results of the St. James Survey, Trinidad. *Int J Epidemiol.* 1989;18:808–817.
13. Emmanuel SC. Trends in coronary heart disease mortality in Singapore. *Singapore Med J.* 1989;30:17–23.
14. Hughes LO. Insulin, Indian origin and ischemic heart disease. *Int J Cardiol.* 1990;26:1–4.
15. McKeigue PM, Shah B, Marmot MG. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *Lancet.* 1991;337:382–386.
16. Seedat YK, Mayet FGH, Khan S, Somers SR, Joubert G. Risk factors for coronary heart disease in the Indians of Durban. *S Afr Med J.* 1990;78:447–454.
17. Ramaiya KL, Swai ABM, McLarty DG, Bhopal RS, Alberti KGMM. Prevalence of diabetes and cardiovascular disease risk factors in Hindu Indian subcommunities in Tanzania. *BMJ.* 1991;303:271–276.
18. Enas E, Yusuf S, Mehta JL. Prevalence of coronary artery disease in Asian Indians. *Am J Cardiol.* 1992;70:945–949.
19. Klatsky AL, Siegelaub AB, Landy C, Friedman GD. Racial patterns of alcoholic beverage use. *Alcohol Clin Exp Res.* 1983;7:373–377.
20. *Manual of the International Classification of Diseases, Injuries and Cause of Death, 1975, Ninth Revision.* Geneva, Switzerland: World Health Organization; 1977.

The Effects of Denturism: New Zealand Dentists' Response to Competition

Nancy Joy Devlin, PhD

ABSTRACT

The New Zealand Dental Act of 1988 allowed clinical dental technicians to deal directly with the public in fitting and supplying dentures. This study tested the hypothesis that dentists responded to competition from dental technicians by lowering their fees. The results indicate that there was no significant change in the fees charged by dentists for dentures. The apparent failure of deregulation to produce the expected outcome could be due to the competitive pressure imposed by dental technicians practicing illegally prior to 1988, to consumers' lack of information, or to barriers to "consumer search" imposed by the act itself. (*Am J Public Health.* 1994;84:1675–1677)

Introduction

Adult dental care in New Zealand is provided mainly by self-employed, privately practicing dentists and is financed predominantly by out-of-pocket payments by consumers; few state subsidies exist and coverage by insurance plans is minimal.[1] In 1988 the New Zealand government changed the legal provisions regarding the practice of dentistry.[2] Regulations that were viewed as conferring "monopoly rights" on dentists were removed, allowing clinical dental technicians ("denturists") to deal directly with the public in supplying and fitting complete and partial dentures ("denturism").

New Zealand legislators expected deregulation to lead to an increase in the supply of dental services, which would, all other relevant variables remaining unchanged, lead to a decrease in the price of these services. It seems plausible to assume that these effects would be evident 4 years after the change in competitive environment. Owing to the use of "grandfather rights" in the registration of previously practicing dental technicians as "clinical dental technicians," the increase in the legal supply of dental services took place within a relatively short time frame. The number of clinical dental technicians registered under the Dental Act was 129 in 1993 (around 4 per 100 000 population); the number of dentists issued annual practicing certificates in that year was approximately 1300. The introduction of the Dental Act increased the total dental work force (those able to deal directly with the public) by around 10%, although the increase in the effective supply of denture services would be larger than 10% because dentists spend a relatively small proportion of their chairside time on dentures (6% in 1988). Fee data from annual surveys of dental practices were used to test the following hypotheses:

1. That the fees charged by dentists for full and partial dentures would increase at a different rate from the fees charged for other services after 1988. As competition from nondentists principally affected the market for dentures, there should be an observable difference between dentists' pricing of dentures and their pricing of other services.

2. That the spread of fees charged by dentists for dentures would narrow in the

The author is with the Economics Department, Otago University, Dunedin, New Zealand.

Requests for reprints should be sent to Nancy Joy Devlin, PhD, Economics Department, Otago University, PO Box 56, Dunedin, New Zealand.

This paper was accepted May 17, 1994.

TABLE 1—Actual and Predicted[a] Fees and Fee Indices for New Zealand Dental Services

	Actual				Predicted, 1992	Difference between Actual and Predicted, %	95% Confidence Interval	r^2
	1978	1983	1988	1992				
Fees[b]								
Examination and simple prophylaxis	$8	$16	$30	$39	$38	+2.6	$34, $44	.974
Single extraction, local anesthetic	$12	$24	$46	$60	$57	+5.3	$70, $50	.962
Amalgam filling, one surface	$7	$15	$30	$40	$37	+8.1	$45, $32	.961
Crown (bonded)	$126	$232	$399	$495	$494	+0.2	$577, $452	.975
Full (upper and lower) dentures	$224	$383	$673	$809	$817	−1.0	$970, $729	.965
Acrylic partial (one tooth)	$62	$117	$203	$259	$262	−1.2	$333, $212	.929
Fee indices[c]								
Denture price index[d]	100	180	314	438	394	+10.0	478, 342	.956
Other dental services[e] price index	100	199	376	551	467	+15.2	558, 431	.968
Relative fee index[f]	100	90	84	80	80	0.0	100, 58	.330

[a]Predicted fees and fee indices for 1992 were estimated from a least squares linear regression of fee data for the years 1978 through 1988.
[b]All fees shown are in current dollar terms (i.e., they are not inflation adjusted) and exclude goods and services taxes.
[c]In each case, the index base is 100 in 1978.
[d]Denture services = partial acrylic, one tooth; complete upper and lower dentures.
[e]Other dental services = examination and simple prophylaxis; single extraction, local anesthetic; amalgam filling, one surface; bonded crown.
[f]Relative fee index = (denture price index/other dental services price index) × 100.

post-1988 period because a wide range of fees charged for the same service is incompatible with the functioning of a competitive market.[3]

Methods

Data on the fees charged by New Zealand dentists are available for 1978, 1981, and annually from 1983 through 1985 and 1987 through 1992, providing a time series spanning 10 years before and 4 years after the 1988 New Zealand Dental Act. These surveys, which are conducted by the New Zealand Dental Association, sample more than 30% of all privately practicing dentists in each year and provide data on the "normal, average, uncomplicated fee" charged by each dentist for a range of items. For the purposes of this study, six fee items were selected: two for dentures (full and partial) and four for other services frequently performed by dentists.

Regression analysis of dentists' fees for the years preceding the Dental Act (1978 through 1988) was used to estimate a least-squares linear trend line, from which expected charges for each item in the period 1989 through 1992 were predicted. These predicted fees were compared with the actual fees charged.

Price indices (with a base of 100 in 1978) were constructed for two groups of dental services—dentures and other dental services—and a relative fee index (RFI) was calculated as follows:

$$\text{RFI} = \frac{\text{Denture Price Index}}{\text{Other Dental Services Index}} \times 100$$

Linear regression of all three indices for the period 1978 through 1988 was used to predict index values for 1989 through 1992, which were compared with actual index values.

Results

Table 1 shows the actual fees charged in selected years and the 1992 fees predicted from pre–Dental Act data. For both full and partial dentures, the actual fees charged in 1992 were lower than predicted; for all other items, 1992 fees were higher than predicted.

Although the direction of these changes appears to support the initial hypothesis, the magnitude of the changes is small. For all treatment items, the actual fee charged in 1992 falls easily within the 95% confidence interval for the 1992 prediction, suggesting that fee charging has not changed substantially for any of these items as a result of the new legislation.

In the post-1988 period, the price index for both dentures and other dental services increased more quickly than predicted (Table 1). The predicted 1992 value of the relative fee index coincided exactly with the actual 1992 value of the index (although the regression r^2 of .33 suggests that no significant linear trend could be fitted). The relative fee index indicates that there was a tendency for denture prices to rise more slowly than other dental prices before 1988, and that this trend continued thereafter. The Dental Act appears to have had no significant influence on the denture price index or the relationship between denture prices and other dental treatment prices.

Table 2 shows the coefficient of variation of fees for each treatment item. There appears to have been little reduction in the dispersion of fees charged for dentures as a result of the Dental Act. The fees charged for full dentures between 1989 and 1992 show the least variability of all treatment items; however, this was also true of the pre–Dental Act period.

Discussion

There are a number of possible explanations for the apparent lack of market response to the Dental Act.

Evidence that the illegal practice of dental technicians was widespread prior to 1988[4] suggests that the main achievement of deregulation may have been to sanction this activity rather than to pro-

TABLE 2—Variability in Fees Charged by New Zealand Dentists, as Measured by the Coefficient of Variation in Fees Charged for Selected Dental Services in 1988 and 1992

Service	Coefficient of Variation[a] 1988	1992
Examination and simple prophylaxis	26	25
Single extraction, local anesthetic	22	22
Amalgam filling, one surface	27	23
Crown (bonded)	19	13
Full (upper and lower) dentures	17	19
Acrylic partial (one tooth)	28	27

[a]Coefficient of variation = (standard deviation/mean) × 100. A high coefficient of variation indicates a greater degree of variability in fee charging between individual dentists.

mote more competition. This possibility is consistent with evidence that the fees charged for full dentures tended to increase less quickly than the fees for other dental services even before the Dental Act. It is also consistent with the finding that the indicator of fee dispersion for full dentures was lower than that for other dental services both before and after 1988.

Legalizing competition may have produced few additional gains because of consumers' lack of information. Whereas denturism has been introduced elsewhere as a consequence of strong consumer lobbying,[5,6] the New Zealand Dental Act was primarily the result of government initiatives. There was little public awareness of, or debate over, the legislation. Consumers' perceptions regarding the quality of care offered by dentists, relative to that offered by technicians, might also partly explain the finding that competition between these suppliers did not affect dentists' fees. Furthermore, advertising by dentists and technicians, although legal, is very low-key in comparison with advertising in the United States, where denturism has been demonstrated to result in lower prices.[7]

The act itself also imposes some barriers to "consumer search"; for example, denturists may fit partial dentures only for persons whose oral health is certified by a dentist.

At the time the Dental Act was introduced, technicians' fees were estimated to be, on average, 30% lower than those of dentists.[8] Dentists' response to the act suggests that these price differentials are likely to persist. □

References

1. Devlin NJ. The dimensions of dentistry: how much do New Zealanders spend on dental care? *N Z Dent J.* 1987;83:66–71.
2. Dental Act No. 150. Wellington, New Zealand: Government Printing Office; 1988.
3. Stigler GJ. The economics of information. *J Polit Econ.* 1961;69:213–225.
4. Newton J. The changing role of dental technicians in New Zealand. *N Z Dent J.* 1989;85:21–22.
5. Rosenstein DI, Joseph LP, Mackenzie LJ, Wyden R. Professional encroachment: a comparison of the emergence of denturists in Canada and Oregon. *Am J Public Health.* 1980;70:614–618.
6. Spencer AJ. Denturism: political and economic factors determining the role of denturism in dentistry. *Aust Soc Prosthet Bull.* 1982;12:7–13.
7. Rosenstein DI, Empey G, Chiodo GT, Phillips D. The effects of denturism on dental prices. *Am J Public Health.* 1985;75:671–672.
8. *Submission to the Social Services Select Committee on the Dentists Bill.* Wellington, New Zealand: New Zealand Institute of Dental Technologists; 1988. (Available from N. Watts, 543 Main Rd, Hatfields Beach, Orewa, New Zealand.)

Public Health Then and Now

The Medical Care Programs of the Farm Security Administration, 1932 through 1947: A Rehearsal for National Health Insurance?

Michael R. Grey, MD, MPH

ABSTRACT

At a time of renewed interest in universal health insurance, an examination of earlier periods when society grappled with the link between socioeconomic status and health is fruitful. Between 1935 and 1947, the federal government sponsored a comprehensive medical care program for low-income farmers, sharecroppers, and migrant workers under the auspices of the Farm Security Administration (FSA). Despite the strong opposition of the American Medical Association, humanitarian and economic concerns at the local level often promoted physicians' participation in the program's group prepayment plans. Many FSA leaders clearly saw the program as a model upon which national health insurance might advance. However, in the wake of World War II, the FSA program declined as physicians' income improved, the rural population declined, and traditional ideological objections to federal intervention in medical care resurfaced. The FSA experience illuminates the complex ideological, economic, and humanitarian motivations of American physicians in the face of health care reform. (*Am J Public Health.* 1994;84: 1678–1687)

Introduction

From 1935 to 1947, the federal government sponsored an extensive civilian medical care program under the aegis of the US Department of Agriculture's Farm Security Administration (FSA). The FSA's mission—to rehabilitate low-income farmers, sharecroppers, and migrant workers—led it to develop a comprehensive medical care program described by the *Saturday Evening Post* as a "gigantic rehearsal for health insurance."[1] At the program's peak, more than 650 000 poor farmers and a million migrants were enrolled in medical care cooperatives or farm labor clinics in a third of all rural counties (Figure 1). Although the New Deal has been richly mined by historians, remarkably little has been written about this "gigantic rehearsal" in the nearly half-century since it ended.[2–4]

Until the passage of Medicare and Medicaid, the FSA program was the largest government-sponsored program dedicated to providing medical care for a specified civilian group. The FSA's success owes much to strategies the agency adopted to promote its medical program among skeptical physicians. These strategies are relevant guidelines as our nation again confronts the issue of national health security. Eager to avoid confrontation with both local physicians and organized medicine, the FSA emphasized free choice of physician and voluntary participation. Its decentralized approach promoted local autonomy and gave physicians substantial but not absolute control over the operation of the medical care plans. Certainly, philanthropies, unions, physicians, and private industry sponsored various prepaid health care plans throughout this period and even earlier. However, the public/private character, extensive enrollment, comprehensive coverage provisions, and preventive orientation of the FSA program gives it a historical import that exceeds that of earlier or parallel health care delivery programs.

Throughout most of this period, the American Medical Association vehemently opposed federal involvement in medical care delivery. In spite of this opposition, physician support of the FSA plans at the grassroots level was substantial and was driven by humanitarian and economic concerns. While physicians saw the program as a temporary federal effort to provide medical care to an indigent group, the agency itself pursued a broader public health agenda. The FSA's extensive public and preventive health efforts and its systematic use of public health nurses, nutritionists, and US Public Health Service medical officers belie the public posture assumed by the agency. Over time, the FSA's multifaceted rural health programs and its eventual alliance with reformers favoring national health insurance made physicians increasingly uncomfortable. This discomfort coincided with physicians' improving incomes and the easing of the economic pressures on them in the years leading up to World War II. Growing congressional opposition to New Deal social legislation, the divisive debate over national health insurance, and concerted opposition to the FSA by conservative farm groups only added to the agency's woes. In sharp contrast, other

The author is with the Department of Medicine and the Department of Community Medicine and Health Care, University of Connecticut School of Medicine, Farmington, Conn.

Requests for reprints should be sent to Michael R. Grey, MD, MPH, Section of Occupational and Environmental Medicine, Bldg 12, University of Connecticut Health Center, Farmington, CT 06030.

privately funded voluntary group prepayment plans (e.g., Kaiser Permanente, Blue Cross, and physician service bureaus) were less vulnerable to attack and made steady gains in the postwar era.[5] In retrospect, however, the root cause of the FSA plans' eventual demise was an ideological conflict between the government and the medical profession.[6,7] For this reason, the history of the FSA medical care program illuminates the ideological, economic, and humanitarian motivations of American physicians in the face of health care reform.

Medicine and Health in the 1930s

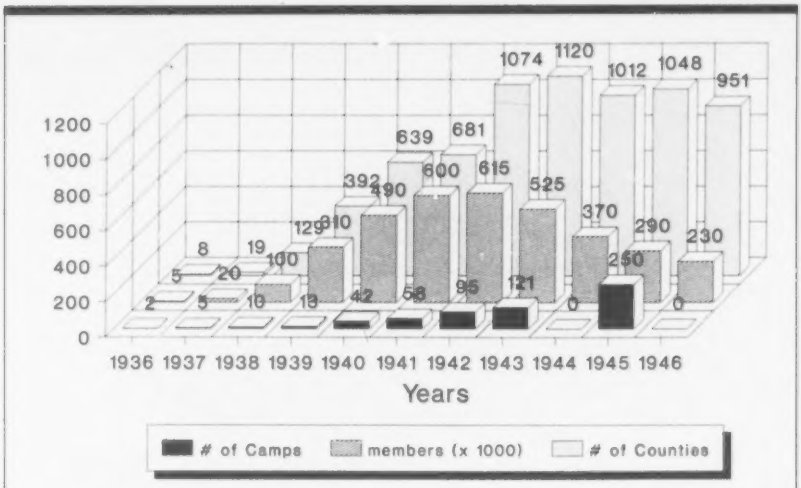
American medicine, like much of society in the 1930s, was in transition. Solo practice and fee-for-service still dominated medical practice, and rural hospitals were few and often proprietary. However, the waning influence of general practitioners, the rising dominance of specialists, and centralization of care in hospitals were well under way by that time.[8] In 1932, the Committee on the Costs of Medical Care published its landmark report *Medical Care for the American People*, the most exhaustive and influential study of the state of American health and medicine that had ever been published. The committee found that poor communities experienced more sickness and received less care than more affluent communities. Medical resources, while plentiful, were not "distributed according to needs, but rather according to the real or supposed ability of patients to pay for services."[9] The numerical, income, and geographical imbalance between general practitioners and specialists led the committee to conclude that the nation needed far fewer specialists and far more general practitioners. Finally, the committee linked access and cost barriers as critical issues for underserved populations, setting the tone for virtually all health care reforms to the present day.[10] An obvious but often neglected fact is that the most nettlesome problems in our health care system antedated changes such as the explosion of medical technology, the acceleration of medical specialization, and the dominance of hospital-based care in the wake of World War II. The committee's 1932 report calling for an integrated system in which generalists provide the majority of acute and preventive services was prescient.

The Great Depression greatly exacerbated but did not create the problems highlighted by the Committee on the Costs of Medical Care. Unemployment soared to an unprecedented 25%, overwhelming private and local relief agencies. Lack of money forced many people to go without medical services, and a single serious illness was enough to plunge a large and steadily increasing percentage of American families into prolonged debt. "I have to treat many families," lamented one physician, "shutting my eyes to the fact that not one of my instructions can be carried out."[11] Physicians' net income plunged by 17%, and many rural physicians' incomes dropped below 50% of billings.[12] Other health care professionals were similarly affected.

The Great Depression devastated rural America. Mechanization and land consolidation, the nation's worst-ever drought, and blunt legislative efforts such as the Agricultural Adjustment Act conspired to create the largest internal migration in our history. Vast numbers of families left their farms buried in dust, loaded up their jalopies, and headed west (see photo on next page). The Department of Agriculture estimated that between 1 and 2 million of the estimated 10.5 million people employed in agriculture were migrants.[13]

The health conditions of rural citizens, which had been declining relative to those of urban Americans since the turn of the century, were also adversely affected. Rural areas had 80 physicians per 100 000 population, vs 171 per 100 000 in urban areas. In 1900, nearly 50% of medical school graduates practiced in rural areas; by 1931, fewer than 21% did so.[14] In metropolitan areas, 72% of births occurred in hospitals, infant mortality was 34.2 per 1000 live births, and immunizations averaged 89%. In contrast, only 14% of rural babies were born in hospitals, rural infant mortality was 43.3 per 1000 live births, and only 37% of rural children were immunized.[15] In New England there were 81 hospital beds per 100 000 population, while in the more rural South there were only 30 beds per 100 000. Ninety-four percent of all water supplies in the South, according to the 1940 census, were open; 66% of Southerners still used privies, and fewer than 12% had potable water within 50 feet. A third of the nation's 3070 counties had no public health unit; virtually all were rural.[16]

Health conditions among migrants were particularly abysmal. Outbreaks of infectious diseases such as typhoid, dysentery, and tuberculosis created vigilante movements that were sometimes led by local health departments. Wrote one county health officer, "One has to deal with a people whose cultural and environmental background is so bad that for a period of more than 300 years no advances have been made in living conditions among them."[17] Racism, xenophobia, and fear of contagion—powerful historical themes in society's response to



Source. Based on data from the BAE report[3(p14)] and Mott and Roemer, *Rural Health and Medical Care.*[2(p407)]

FIGURE 1—Evolution of Farm Security Administration health programs, 1936 through 1946.

"Mother and Daughter Living on the Side of the Road," Nipoma, Calif, 1936. Photograph by Dorothea Lange, from the Library of Congress, Prints and Photographs Division, FSA Historical Division (LC-USF34-33601 17521).

disease epidemics—led to the violent and systematic destruction of squatter camps or "Hoovervilles." To many people, disease and degraded mortality seemed equally contagious.[18]

Roosevelt quickly moved to promote massive federal intervention during the famous "100 days" of his presidency. Congress passed an omnibus relief measure creating the Federal Emergency Relief Administration (FERA) in March 1933. The FERA channeled direct federal relief through state emergency relief administrations and created a division devoted solely to rural relief and rehabilitation. The hallmarks of the FERA rural rehabilitation program—friendly supervision and easy credit—remained at the core of all subsequent efforts. The creation of the Works Progress Administration (WPA) and the Resettlement Administration in 1935 signaled a shift in federal policy away from direct monetary relief. The more well-known WPA concentrated on massive infrastructure projects and urban work relief. The Resettlement Administration assumed the rural rehabilitation prerogatives of the FERA and operated for 2 years as an independent cabinet-level agency under brain truster and political lightning rod Rexford Tugwell. In 1937, the president renamed the Resettlement Administration the Farm Security Administration and placed it in the more conservative Department of Agriculture. In 1943, as part of wartime restructuring, the War Food Administration assumed responsibility for the FSA migrant programs.[19] (For clarity, the acronym FSA is used throughout this essay.)

The FSA promoted marketing, farming, and equipment-buying farm cooperatives to help smaller and poorer producers compete in the agricultural marketplace. The agency also believed that these cooperatives would promote economic stability, enhance self-reliance, and foster local leadership. Pressure on the agency to maintain good loan repayment among its rehabilitation clients soon conflicted with the FSA's humanitarian thrust, and when it became clear that ill health was responsible for 50% of all loan defaults, the FSA moved into the field of health care delivery. As US Public Health Service senior surgeon and FSA chief medical officer Ralph C. Williams stated, "a family in good health was a better credit risk than a family in bad health."[20]

Medical Care Cooperatives

Williams told those attending the 1939 American Public Health Association convention that the FSA medical program was an "incidental by-product of a depression-born loan program for farm families unable to obtain credit elsewhere, and designed to accommodate a special economic group only."[21] This economic justification pacified vocal groups unsympathetic to the agency's social agenda, such as organized medicine, conservative politicians, and organized farm groups. However, Williams' public posture understated the powerful ideological commitment of the agency's medical hierarchy to make more public the practice of medicine. In line with the agency's cooperative philosophy, local FSA supervisors encouraged farmers to establish medical cooperatives. These supervisors asked local physicians to provide care to FSA clients in a group prepayment scheme to lower cost barriers and ensure access to needed medical care. Bundled into their annual loans, rehabilitation clients (also called borrowers) received a federal subsidy (typically around $35), which they then paid into a trustee-supervised fund. Participating physicians billed this fund, and if billings exceeded the amount set aside that month, doctors received prorated reimbursement. Flexibility at the local level was critical to the program's success with farmers and physicians alike. The policy of promoting cooperatives and local determination of local needs also fit into the FSA's commitment to participatory democracy. "Any plan in which the families unite to help themselves should reduce the cost of medical aid and thereby make more effective the funds thus expended [so that] a worthwhile beginning can be made by the families themselves toward better health."[22] Healthier clients made the task of supervision easier, and credit risks diminished as clients' health improved. This fact not only was a source of pride for the agency but was also critical to continued congressional support. Nearly 90% of all loans were eventually repaid in full. The FSA plans were also a new source of income for hard-pressed rural doctors. Participating doctors collected 65% of their fees from a group that had previously been able to pay little, if anything, for medical care.[23]

It was not the intent of the FSA in the medical cooperative program to fundamentally restructure the delivery of rural health services. Group prepayment was grafted onto traditional fee-for-service practice. Concessions of this sort, along with the fact that physicians' participation was voluntary and the extension of care was limited to a specified low-income group, made the FSA programs palatable to financially strapped rural practitioners. Still, the agency's promotion of consumer

participation on the cooperatives' governing boards, the key fiscal and administrative role played by the FSA, and the lowering of cost and access barriers for low-income farmers represent genuine reforms for the era. Of equal importance, the medical care cooperative program gave the FSA experience it put to good use as it extended its health programs to migrant workers and experimented with more radical proposals.

Migrant Health Programs

The New Deal confronted the problem of migrant labor in the United States in early 1935 when the California state emergency relief administration built two farm labor camps, using federal and state rural rehabilitation funds. The mobility of the migrants made direct payments for medical care administratively cumbersome. Consequently, in 1937 the FSA devised a bold alternative means of providing medical services: Agricultural Workers Health Associations (AWHAs). By 1946, seven regional AWHAs were providing medical services in 250 federal migrant camps nationwide.

The FSA developed its first AWHA in California and Arizona following a tour of the region by Dr Ralph C. Williams. Shocked at the appalling health conditions of migrant families, Williams called for the creation of a "state-wide incorporated health association . . . to make direct payments to doctors and hospitals for medical and hospital treatment."[24] Williams wisely enlisted the support of leaders in the California Medical Association. In 1937, the FSA gave $1 million to the newly incorporated Agricultural Workers Health and Medical Association. California and Arizona physicians made their uneasy alliance with the government, mindful of the swelling tide of social and political support for medical care reform:

> There have been theoretical objections and charges that such plans are a step toward "State Medicine," "compulsory sickness insurance" or "socialized medicine." In evaluating such objections, it may be well to consider alternatives. News . . . reveals almost parallel movements . . . in the direction of greater intervention of government in medical care. [S]tatements of President Roosevelt and the Social Security Board officials [and] the effort of some hospital insurance plans to invade the field of medical practice . . . are portents of a trend in the United States. It is doubtful if head on opposition can greatly affect this trend.[25]

As was true with the medical care cooperatives, local FSA representatives, typically camp managers, recruited physicians into the AWHAs. AWHA doctors rendered services at agreed-upon fees and the AWHA paid them directly. Physicians' participation was voluntary, and the physicians insisted that the migrants have free choice in selecting their physicians. Full-time public health nurses staffed the camp clinics and referred patients to participating physicians' offices when necessary. Physicians also held clinic hours at specified times of the week and received payment on a per diem, hourly, or—occasionally—salaried basis. Whenever possible, the AWHAs used existing community hospitals as referral centers for their sickest migrants. At Eleven Mile Corner, Ariz, and Belle Glade, Fla, the FSA took the extraordinary step of owning and operating hospitals for its migrant clients.[26]

FSA district and regional medical officers showed pragmatism and flexibility in adapting the AWHAs to local needs and exigencies. Medical societies influenced but did not dictate the setting of fee schedules and limits of coverage, and they could even express grievances on staffing matters to FSA medical officers. For their part, FSA district and regional medical officers maintained liaisons with county and state medical societies, supervised the field medical and nursing personnel, and implemented public health policies whose influence often extended beyond the camps' perimeters. Considerable local autonomy resulted in substantial variations in the organization and character of the migrant health programs. In California, migrants were mainstreamed into local physicians' offices, while elsewhere camp medical clinics were the preferred sites for medical care. FSA programs occasionally became sources of care for nonmigrant FSA rehabilitation clients. Despite local variations, the structure and philosophy of the AWHAs were consistent from region to region: available and adequate medical services, government subsidization of costs, and a focus on preventive services, particularly for pregnant women and children.

By negotiating fee schedules, granting physicians substantial freedom in medical decision making, and making the programs voluntary, the FSA successfully used a strategy that was repeated in many subsequent battles over health care reform. Even more than its medical care cooperative program, the FSA's promotion of an organizational and fiscal entity such as the AWHA to provide access to health care signified a major shift in federal health policy. Using the AWHA as a surrogate payer preserved the appearance of independence upon which doctors insisted while effectively undermining physicians' opposition to more overt government involvement in medical care financing or delivery.

Resettlement Projects

One of the FSA's more controversial efforts was the creation of a series of resettlement projects and greenbelt communities throughout the country (e.g., Greenbelt, Md). Under the resettlement program, low-income farmers, sharecroppers, and farm laborers living in the resettlement "homesteads" were given housing, a small plot of land for personal use, and technical advice on cooperative farming and marketing. The federal government's promotion of the resettlement projects embroiled the FSA in a political maelstrom from which it never fully extricated itself. Critics charged that land redistribution and agricultural cooperative programs were proof that the FSA had been "vaccinated by the Tugwellian virus [and] placed into the hands and under the blight of social gainers, do-gooders, bleeding hearts, and long-hairs who make a career of helping others for a price and according to their own particular, screwball ideas."[27]

Resettlement projects provided on-site care by general practitioners, although there was a wide mix of hospital, dental, and specialist care. The FSA established health clinics staffed by full-time salaried nurses and either full-time or part-time salaried physicians in at least 40 resettlement projects. Often, resettlement communities formed incorporated health associations similar to the AWHAs and contracted with local physicians, hospitals, and dentists for discounted services. Archival evidence, although anecdotal, provides support for the benefits of the program. In 1939, the FSA's chief medical officer found that "the families require less care now than they did 2 years ago, that they abuse their privileges very little, that they call him early in illness, that he enjoys making repeat calls without feeling it a burden on the families, and that the general Farm Security Administration program has already shown results in healthier families, particularly the children."[28]

Agricultural Workers Health Association clinic, Bridgeton, NJ, 1942. Photograph by John Collier, from the Library of Congress, Prints and Photographs Division, FSA Historical Division (LC-USF 34-83435 17521).

Statewide Prepaid Plans

Perhaps no other region suffered the effects of the devastating drought of the 1930s more than North Dakota and South Dakota, two states where virtually everyone qualified for relief. In August 1936, alarm over rapidly deteriorating health conditions in the region prompted a meeting of local, state, and federal officials, including Surgeon General Thomas Parran, the WPA, the US Children's Bureau, the FSA, and organized medical groups of both states. Within a year, the FSA established two statewide capitated medical care plans in the drought-stricken Dakotas. The North Dakota and South Dakota Mutual Aid Corporations provided home and office medical care, hospitalization, major surgery, emergency dental care, nursing care, and drugs. The fee schedule was nearly 30% less than the prevailing minimums. Prevention and maternal and child services were integral components of these statewide programs, and the Mutual Aid Corporations even promoted clinical practice guidelines such as "an agreed minimum number of prenatal visits, delivery in the home or hospital, and necessary post-natal care."[29] Despite the fact that they received 60% of their billings, physicians opposed the prorating of bills so vigorously that the two statewide medical care experiments ended after 2 years of stormy operation.[30]

Experimental Rural Health Programs

Although the medical care programs and the AWHAs represented the most extensive and successful of all FSA medical care efforts, the agency undertook demonstration projects that were far more controversial. These experimental health programs went significantly beyond what agency leaders considered to be limitations of the earlier and more established programs for rehabilitation borrowers and migrant workers. In these earlier efforts, voluntary enrollment—insisted upon by physicians—led to an adverse selection effect. Families most likely to use services enrolled in the group payment programs while healthier families preferentially opted out of the group payment plans in favor of traditional fee-for-service arrangements. Certainly, pressure from local physicians strongly influenced this trend. Yet despite strenuous educational efforts by the FSA, the economic advantage of paying in advance for medical care, regardless of anticipated need, remained a difficult concept for many farmers. In response to this trend, the FSA considered alternative ways of improving the actuarial soundness of group medical care plans. The FSA developed experimental rural health programs to resolve this dilemma. Fiscal stability was achievable, the agency argued, by distributing the costs of illness over a broader patient base through countywide medical care plans with no income restrictions. Agency leaders also believed this strategy could promote higher reimbursements for physicians while allowing yearly membership fees to be income-adjusted and moderate, thereby generating broader community support.

Beginning in 1942, the FSA selected seven rural counties with an interest in medical care, reasonably stable income, sufficient medical and hospital facilities, and receptive local physicians. Of cardinal importance was the FSA's preference for counties with full-time public health units. This criterion was a reflection of the long-standing relationship between the FSA and the Public Health Service. Indeed, many of the leading figures in the medical care experiments of the FSA were Public Health Service officers who shared that agency's commitment to bringing public health back into the heart of medical care delivery. The experimental health programs adopted a sliding-scale membership fee, and the government subsidized the remaining costs. This federal support was necessary to enable lower-income families to participate and to allow the programs to include preventive services, hospitalization, drug costs, and dental work. The FSA created six rural health services (as they were generally called) in Texas (Wheeler and Cass counties), Mississippi (Newton County), Nebraska (Hamilton County), Arkansas (Nevada County), Georgia (Walton County), and New Mexico (Taos County). An even more ambitious six-county experimental health program was begun in southeastern Missouri.

The services provided by the experimental health programs reflected the FSA's consistent public health and primary care orientation. General practitioner services included office, home, and hospital visits. Surgical or subspecialty consultation was available with a referral by the general practitioner. Dental services covered not only emergency care but also routine and prophylactic treatment. All seven programs initially covered drug expenses. The experimental health programs maintained the characteristic flex-

ibility of the FSA medical care programs. Membership fees based on ability to pay, voluntary choice of available physicians, broad scope of services, voluntary participation by physicians and farmers alike, and local administration were constants. However, financing mechanisms varied and included modified fee-for-service plans, capitation, and full- or part-time salaried medical staff.[31]

Acceptance of these experimental programs was generally high among rural families eligible for participation, and membership as a percentage of population in the first year of operation ranged from 32% to 74%. Turnover remained a serious problem, however, sometimes reaching 50% per year.[32] When queried about their lack of ongoing interest, farmers usually cited the high membership costs of the program and their feeling that they were not receiving the services for which they had paid. Members were often unaware that the federal government subsidized most of the programs' costs, and the FSA found that out-of-pocket payments for medical care, alternative healers, and drugs were common. Although the experimental health programs were explicitly designed for a broad cross-section of the community, the public's perception that they were for disadvantaged farmers was hard to eradicate. Doctors reinforced this perception and discouraged middle-income families from joining.[33]

Few credible data on whether the experimental health programs had a beneficial effect on members' health exist. Basic vital statistics at that time were notoriously inaccurate, particularly in rural areas. It was also an era when few questioned the assumption that more medical care resulted in improved health:

> More mothers are now receiving prenatal and postnatal care, and better medical care and more hospitalization at childbirth. Patients are going to a doctor oftener and earlier, and are making more use of hospitalization facilities. Many families reported an increased feeling of security.[34]

Viewed in this light, the experimental health programs and the FSA program in toto were successful by commonly accepted measures, such as number of physician visits, hospitalization rates, number of prenatal visits, number of physician-attended or hospital births, and immunization rates. Experimental health program members received, on average, 2.6 physician visits per year, double the number cited by the Committee on the Costs of Medical Care for rural citizens. Prior to the establishment of the Nevada County program, only 10% of White women in the county delivered in hospitals, although nearly 90% were attended by a physician. In sharp contrast, Blacks delivered at home 99% of the time, nearly all attended by midwives or experienced neighbors. After the experimental program began, 50% of White babies and more than 25% of Black babies were born in the local hospital, while the percentage of physician-attended births for Blacks climbed to 75%.[35] Similar dramatic changes in obstetrical practice occurred in Taos County, which had a predominantly Spanish-speaking population. The annual number of hospitalizations for delivery in the fledgling Blue Cross plans was 107 per 1000 members, compared with 155 per 1000 in the experimental health programs.[36]

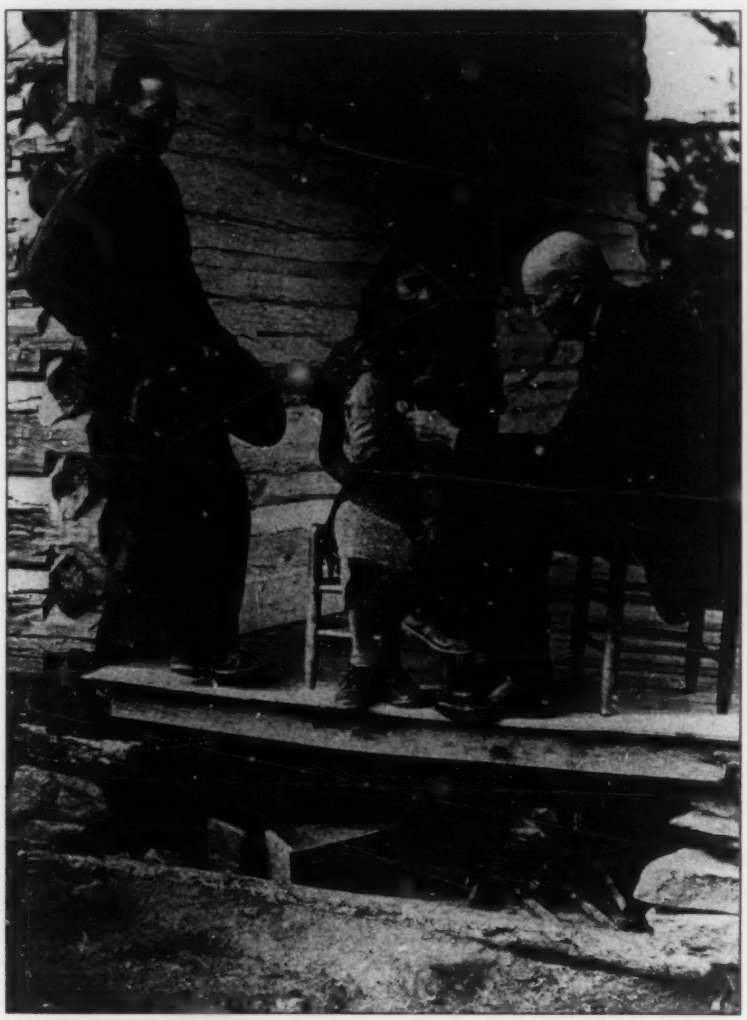
The experimental health programs were born of the conviction of FSA leaders that major structural and financial changes in America's health care system were imminent. A 1944 poll found that 68% of Americans supported an extension of Social Security to cover medical care and hospitalization; 89% felt that some people did not get medical care because of cost barriers; and 33% had themselves deferred treatment for financial reasons.[37] Many ranking FSA medical officers firmly believed that the nation would eventually adopt some form of national health insurance. These leaders believed that the experimental health programs afforded an unparalleled opportunity to gain practical experience in operating what might become a model for a future national system of medical care delivery. Former FSA chief medical officer Frederick D. Mott and his former assistant Milton I. Roemer wrote in 1948 that the experimental program had "given us what is probably an unmatched body of experience in the possible achievements and deficiencies of the tax-assisted and voluntary health insurance plans in rural areas."[38]

The experimental health programs experienced many of the same problems that beset earlier FSA health care delivery programs. The perception that the programs were for the indigent, reinforced by physicians who discouraged higher-income families from signing up, and the high turnover rate among families who used fewer services adversely affected the programs' actuarial soundness. The drain of younger, more able-bodied rural citizens to the armed services and urban-based defense industries left behind groups—women, children, and the elderly—with higher than average health care utilization, further exacerbating the adverse selection effect. Although professional misgivings were evident early on, many physicians supported the plans with the understanding that "the proposed association would in no way interfere with physicians in the practice of medicine and that patients would have free choice of physicians."[39]

The Medical Profession's Response

The reasons behind the FSA's initial success in recruiting rural practitioners are worth examining. First, the FSA's explicit goal was to provide medical care for a low-income population that was severely straining local medical resources. While doctors worried about the potential of the FSA programs in terms of a broader federal health agenda, the plans provided timely financial relief to hard-pressed rural general practitioners, driving them into a temporary alliance despite opposition by the American Medical Association. Free choice of physician, voluntary participation by client and practitioner alike, and control over matters of medical practice were core elements in the FSA health program. The FSA dealt only with medical doctors or their representatives, excluding osteopaths, chiropractors, midwives, and other folk practitioners. This policy strengthened the government's hand with the highly vocal and well-organized medical profession, the group whose authority over the practice of medicine was approved of by a public increasingly convinced of the benefits of scientific medicine.

At the center of the FSA plans were primary care practitioners, both physicians and nurses. This emphasis was crucial to the programs' successes for more than a decade. While not without its pragmatic side (after all, the majority of physicians available to the agency in rural areas were general practitioners), this strategy mirrored the convictions of the FSA medical leadership and harkened back to the recommendations of the Committee on the Costs of Medical Care. Well before World War II, the complementary forces of specialization and centralization were undercutting the central presence of the general practitioner in American medicine. Thus, the FSA programs bolstered the flagging financial and professional fortunes of the rural general-



Caswell County, North Carolina, 1940. Photograph by Marion Post Wolcott, from the Library of Congress, Prints and Photographs Division, FSA Historical Division (LC-USF 34-56376-D).

ist and provided doctors with a powerful, if potentially worrisome, ally in the profession's emerging sovereignty over medical practice.

Local, regional, and national politics influenced physicians' receptivity to the FSA plans. Statements that the plans violated medical ethics and the sanctity of the physician–patient relationship were commonplace. Many county and state medical societies, and invariably the American Medical Association, preferred the status quo to any government-sponsored health insurance program. As late as 1938, editors of many state medical journals mouthed official American Medical Association policy: "These are problems for local and state consideration . . . rather than problems of federal responsibility."[40] While physicians who supported the plans defended their position in the *Journal of the American Medical Association* and state medical journals, many doctors vociferously contended that the plans provided inferior care and that "medical services did not lend [themselves] to cooperative handling."[41] Clearly, medical care plans based on anything but the traditional fee-for-service engendered significant distrust among physicians. Even at the county level, where support for the FSA programs was most deep rooted, conflicts invariably emerged. Physicians charged above the established fee schedule, collected fees on the side, and double billed—all in violation of their agreements with the FSA. Physicians pressured rehabilitated families to cancel their membership, since a family that had paid off its debt was "no longer in need and [could] afford to purchase medical services privately."[42] Finally, many physicians responded to the prorating of bills by limiting service rather than increasing their attention to prevention, as the agency had naively expected. Doctors preferred strict definitions of medical eligibility, while the FSA's eligibility criteria and definitions of medical need were generous. For physicians, the laxity of the FSA's eligibility criteria, exemplified by the use of the AWHA camp medical clinics by nonindigent local residents, was an omen of a more threatening, unstated federal agenda.

Tension between the government's advocacy on behalf of low-income rural Americans and the medical profession's desire to control the practice of medicine eventually threatened the uneasy alliance that promoted the federal programs. Conflicts over the composition of supervisory boards, eligibility criteria, and, of course, payment were continuous. World War II's effect on already scarce rural medical manpower, the shifting funding priorities of Congress, and a growing political backlash against the social agenda of the New Deal exacerbated many of these conflicts. The impressive growth of the FSA health program notwithstanding, difficulties were evident even in the earliest stages of the FSA's involvement in health care. From 1938 on, termination of medical cooperatives accelerated, although the program peaked in terms of membership in 1942.[43] Powerful and deeply felt ideological positions within the medical profession had as much to do with the eventual dismantling of the medical care programs as the disruption caused by the war.[44]

The Great Depression marked a significant shift in the locus of charity and public welfare programs from philanthropists and local government to the federal government. The 1935 Social Security Act and other New Deal social legislation irreversibly broadened the federal government's responsibility for social welfare in America. While the American Medical Association stressed throughout the Depression that the care of indigents was a local concern, powerful financial and humanitarian concerns temporarily over-

rode physicians' traditional ideological opposition. Many local physicians and state medical societies accepted the FSA's role as a fiscal, organizational, and administrative intermediary. This schism certainly promoted the remarkable growth and development of the FSA's health programs. The threat of national health insurance drove constituent medical societies and individual practitioners to consider alternative methods of financing and delivering medical care.

The suspicions voiced by physicians that the FSA programs were a harbinger of a federalized health care system were not mere professional paranoia, nor were they inconsistent with many readily apparent facts. FSA leaders anticipated fundamental changes in the nation's health care system, and many in Washington expected the vast experience gained in the FSA plans to shape these reforms. FSA leaders actively participated in the deliberations of the Interdepartmental Committee for Health and Welfare Activities that supported the expansion of public health services and grants-in-aid to states for hospital construction, programs for the medically needy, and state-based general medical care programs. The committee's recommendations led directly to the 1938 National Health Conference. Soon after this conference, the American Medical Association reversed its long-standing opposition to voluntary health insurance in an extraordinary emergency meeting of its house of delegates.[45] In 1942 the FSA took a leading role in the Interbureau Committee for Post-War Programs in Agriculture, charged by the president to develop "all possible phases of a health program for the entire farm population now."[46] The FSA's ownership of two hospitals, its promotion of salaried and clinic-based practices, and its experimental health programs testify to agency leaders' interest in pushing the boundaries of federal involvement in health care delivery. The frequent use of Public Health Service officers to provide, administer, and organize rural medical services—an early harbinger of the National Health Service Corps—is further evidence of this agenda. In 1946 a youthful New Deal senator from Florida, Claude Pepper, noted in particular that the experimental health plans, "based on the tax-assisted voluntary health association principle, constitute a series of experiments of interest to the whole Nation. They are particularly important at this time, when our whole future national health policy is being decided."[47] Dr Mott confirmed Senator Pepper's views, saying that the experimental health plans "hold lessons for every rural community and for urban as well as rural Americans . . . [a]t a time when the nation is facing its health problems, weighing possible solutions, and rapidly approaching the stage of long-needed action."[48] It was with conscious deliberation that the agency downplayed this more progressive agenda during the evolutionary phase of the medical care plans.

Events in Europe and increasing fiscal and political conservatism in Congress profoundly altered the reform spirit characteristic of the early Roosevelt presidency. Between 1942 and 1946, the FSA embarked on a number of creative efforts to stave off criticism, such as having its nurses conduct Red Cross first aid programs. These efforts ultimately proved unsuccessful. The FSA was unable to withstand the attacks of conservative farm organizations and congressional opponents who had long waited to eviscerate the agency. In 1947, the FSA was transmutated into the far meeker Farmers Home Administration. Simultaneous with these developments, the allegiance of physicians shifted to groups antagonistic to the FSA's reform agenda. Dr Mott's warning in a confidential memorandum to his staff that an "unholy alliance" between the American Medical Association and the Farm Bureau could undermine the FSA health programs proved valid.[49] For example, in 1946 the California Medical Association abrogated its agreement with the FSA in favor of a group insurance program sponsored by the Farm Bureau and the Grange, two groups whose antipathy to the FSA was well known.[50]

Factors operating on a national scale were catalysts in the decline of the FSA's political fortunes. The war economy's beneficial impact on physicians' income and the scarcity of medical personnel resulting from the draft lessened the economic incentives that had previously generated flexibility on the part of doctors.[51] Stiffening resistance to further federal encroachment into medical care delivery was evident in physicians' initial opposition to the Emergency Maternal, Infant and Child Act.[52] The stumbling momentum toward national health insurance further crippled the agency, whose medical leaders were by then openly in support of a comprehensive national health program. The difficulties faced by the FSA medical leadership since the program's inception were evident in Dr Mott's response to a request to present his agency's position on the 1943 Wagner-Murray-Dingell national health insurance bill. The request, wrote Dr Mott confidentially to his colleagues, "raises an important question about the dilemma—which is not new to any of us—that of working for certain goals privately and yet having to take carefully considered stands publicly."[53] Indeed, as a matter of political and pragmatic expediency, FSA medical leaders downplayed their support of national health insurance. Privately, they remained optimistic that the FSA experience would move a national health program forward, although their optimism went unrewarded.[54]

Conclusion

The experience of the FSA health program yields insights into the often uncomfortable relationship between medicine and the federal government during periods of health care debate. Historically, the medical profession has been loath to place itself at the forefront of health care reforms. However, under financial duress, carried in the wake of a nationwide spirit of social reform, and undoubtedly faced with interventions that appeared inevitable, physicians played a vital role in the FSA health care program. Left to their own impulses and freed from the coercive effect of economic distress, they renewed their traditional opposition to federal involvement in health care. Although the FSA program expanded until World War II, physicians' dissatisfaction with and nonrenewal or outright abrogation of agreements accelerated from 1938 forward. Ultimately, physicians remained wary of third-party intermediaries and willingly allied themselves with the organized farm groups who were the architects of the FSA's demise.

The federal government's foray into health care delivery was not a wasted effort. By intervening directly in the provision of medical care services for underserved civilian populations, the government established a broad precedent. Not until the passage of Medicare and Medicaid, the creation of the Office of Economic Opportunity, and the neighborhood health center movement during the Kennedy-Johnson era would such a vast number of poor Americans gain access to medical care through lowered cost barriers. The economic and humanitarian factors that promoted the FSA's program are of central importance. Certainly, hundreds of thousands of rural Americans experienced their first regular source

of medical care and received immunizations, nutritional advice, and physical nurture within the friendly, if paternalistic, embrace of the federal government. By participating in FSA plans, physicians served both their financial and their more noble goals. Both physicians and clients experienced firsthand the role that fiscal and administrative intermediaries could play in the traditionally sacrosanct doctor–patient relationship. Postwar history shows that it was the private sector and not the federal government that benefited most from this initiation process.[55]

The agency's interdisciplinary approach to providing services demonstrates its public health and primary care orientation. The FSA's plans were centrally financed but organizationally decentralized. Since they revolved around the provision of care by generalists, they ran counter to the prevailing direction of American medicine. The plans reinforced the professional and economic security of rural general practitioners. The FSA's routine incorporation of such supposedly recent health care innovations as the primary care gatekeeper, the use of nurses as physician extenders, and preventive service was also a notable accomplishment and reflects an early federal experiment in managed care. Finally, many at the time believed that the FSA experience would shape a national health insurance plan that they believed was imminent. Congressional hearings on national health policy and the visible and active role given to the FSA medical hierarchy in postwar health care planning substantiate this view.

Although it was not the focus of this essay, the FSA's influence rippled well beyond the 12 years of the agency's existence. Many public health–oriented physicians who worked in the FSA program went on to shape future health care reforms, both in the United States and abroad. Dr Mott moved to Saskatchewan in 1945 and helped mold its provincial health care plan, which was the model for Canada's national health care system. The progressive movement in medicine and public health, programs such as the United Mine Workers Health and Welfare Fund, and the Office of Economic Opportunity owe a debt to such luminaries as Milton Roemer, Loren Kerr, George Silver, Sy Axelrod, Les Falk, Joseph Mountin, Martha May Eliot, Henry Makover, and many others with ties to the FSA.[56] As Dr Mott wrote to his dispirited colleagues from his vantage point in Canada, "We can have the satisfaction of spearheading a movement which will ultimately benefit every citizen. . . . [O]ur consolation can be that we played an historic role at a time when pioneering was urgently needed."[57]

The FSA medical care program remains one of the earliest, most extensive, and certainly most comprehensive federal efforts in health care delivery. It incorporated precepts usually considered more contemporary, such as consumer participation, decentralization of care, centralization of payment source, and an abiding focus on prevention and health education. Strategies that controlled costs, such as the use of nurse clinicians, salaried physicians, and administrative controls on referrals and hospitalization, all speak to the agency's prescient approach to health care delivery. Finally, the FSA provided hundreds of thousands of rural families with their first taste of insurance-based medical care. One participant believes that the FSA experience helped spur the dramatic postwar growth of the private health insurance industry in rural America.[58]

The FSA experience illuminates many characteristics of American medicine that are relevant to the current health care policy debate. Its historical importance as the most extensive New Deal foray into health care delivery notwithstanding, the experiment foreshadowed or directly influenced various health care delivery models in this country and abroad. Forces that contributed to the FSA's remarkable growth and eventual demise provide insight into the role of government and physicians in health care reform. Universal health insurance has resurfaced on our nation's political agenda. The FSA experience, seen as a dry run for a national health program, undermines the myth that government cannot play a creative, fiscally responsible, and professionally acceptable role in health care financing and delivery. □

Acknowledgments

Research for this article was supported in part by Smith Kline Beckman Creative Perspectives in Medicine, the Robert Wood Johnson Clinical Scholars Program, and the National Endowment for the Humanities.

Portions of this paper were presented at the annual conferences of the Society of General Internal Medicine, the American Public Health Association, and the Robert Wood Johnson Clinical Scholars Program.

I want to thank my colleagues in the Department of Medicine, especially Drs Eileen Storey, Michael Hodgson, Marcia Trape, and Henry Schneidermann, for their review of earlier drafts. My research assistant Tarak Kundestuncer was helpful in preparing the final manuscript.

References

1. S. Lubell and W. Everett, "Rehearsal for State Medicine," *Saturday Evening Post*, December 17, 1938.
2. References to the Farm Security Administration (FSA) can be found in Rickey Hendricks, *A Model for National Health Care: The History of Kaiser Permanente* (Brunswick, N.J.: Rutgers University Press, 1993), 14, 18; Paul Starr, *The Social Transformation of American Medicine* (New York, N.Y.: Basic Books, Inc., 1978), 271; Rosemary Stevens, *American Medicine and the Public Interest* (New Haven, Conn.: Yale University Press, 1974), 288n. A longer but not particularly analytical discussion can be found in Bess Furman, *A Profile of the U.S. Public Health Service, 1798–1948*, GPO publication NIH 73-369 (Washington, D.C.: US Government Printing Office, 1973), 401–404. The most comprehensive description of the FSA's plans was published in the immediate aftermath of the program's demise by two senior medical leaders within the FSA. See Frederick D. Mott and Milton I. Roemer, *Rural Health and Medical Care* (New York, N.Y.: McGraw-Hill Publishers, Inc., 1948).
3. Michael R. Grey, "Dustbowls, Disease, and the New Deal: The Farm Security Administration Migrant Health Programs, 1935–1947," *Journal of the History of Medicine and Allied Sciences* 48(1993):3–39.
4. Michael R. Grey, "Poverty, Politics, and Health: the Farm Security Administration Medical Care Programs, 1935–1945," *Journal of the History of Medicine and Allied Sciences* 44(1989):320–350.
5. A recent comprehensive history of the Kaiser Permanente programs is provided in Hendricks, *A Model for National Health Care*. A less detailed discussion of the growth of private insurance plans and of the physician service bureau movement can be found in Stevens, *American Medicine and the Public Interest*, 267–289, and Starr, *The Social Transformation*, 290–334.
6. Grey, "Dustbowls, Disease, and the New Deal."
7. Grey, "Poverty, Politics, and Health."
8. See Stevens, *American Medicine and the Public Interest*, 293–317. A full account of these developments appears in Rosemary Stevens, *In Sickness and in Wealth: American Hospitals in the Twentieth Century* (New York, N.Y.: Basic Books, Inc., 1989).
9. Committee on the Costs of Medical Care, *Medical Care for the American People: The Final Report of the Committee on the Costs of Medical Care*, publication no. 28 (Chicago, Ill.: University of Chicago Press, 1932), 4–5.
10. I. S. Falk, "Some Lessons from the Fifty Years since the CCMC Final Report, 1932," *Journal of Public Health Policy* 4(1983):135–161.
11. Clarence Enzler, *Some Social Aspects of the Depression: 1930–1935* (Washington, D.C.: Catholic University of America Press, 1939), 42.
12. Richard Hellman, "The Farmers Try Group Medicine," *Harper's Magazine*, December 1940, 7.

13. These statistics were excerpted from the following papers located in the personal records of Henry C. Daniels: U.S. Dept. of Agriculture, Bureau of Agricultural Economics and Farm Security Administration, *Background of the War Farm Labor Problem* (Washington, D.C.: US Government Printing Office, May 1942), 15–21; Henry C. Daniels, *History of the Farm Security Administration Camp Program for Migratory Agricultural Workers* (Washington, D.C.: U.S. Dept. of Agriculture, Farm Security Administration, 1944), 15.
14. Frederick D. Mott, *Health Provisions in a Social Security Program for Farm People, with Special Reference to the Experience of the Health Programs of the Farm Security Administration* (Washington, D.C.: U.S. Government Printing Office, September 1944), 1–44. File 31-2, Frederick Dodge Mott papers, Canadian National Archives, Ottawa, Canada (hereafter cited as Mott papers).
15. "National Conference on Rural Health," *Journal of the American Medical Association,* organizational sections, 131(1946):553–561.
16. Mott, *Health Provisions.*
17. Walter J. Stein, *California and the Dustbowl Migration* (Westport, Conn.: Greenwood Press, Inc., 1973), 60.
18. Social historians have traced this theme with regard to a number of public health menaces. Excellent sources are Charles E. Rosenberg, *In The Cholera Years* (Chicago, Ill.: University of Chicago Press, 1962); Rene Dubos and Jean Dubos, *The White Plague: Tuberculosis, Man, and Society* (Boston, Mass.: Little, Brown & Co., 1952; reprint, with new foreword and introduction, New Brunswick, N.J.: Rutgers University Press, 1987); and Allan M. Brandt, *No Magic Bullet: A Social History of Venereal Disease in the United States since 1880* (New York, N.Y.: Oxford University Press, 1987).
19. The only comprehensive institutional history of the FSA is Sidney Baldwin, *Poverty and Politics: The Rise and Decline of the Farm Security Administration* (Chapel Hill, N.C.: University of North Carolina Press, 1968).
20. U.S. Dept. of Agriculture, Farm Security Administration, *Medical Care for FSA Borrowers* (Washington, D.C.: U.S. Government Printing Office, June 15, 1939), 2. Records of the Farm Security Administration, Record Group 96, National Archives, Washington, D.C.
21. Ralph C. Williams, digest of a speech delivered to the American Public Health Association, November 1, 1939. Mott papers.
22. Maguire to all regional directors, May 4, 1938; Health and Rehabilitation file, Regional Medical Programs, 1937–1939; correspondence of James S. Heizer, executive assistant. Records of the Farm Security Administration, Region IV, Office of the Director, Record Group 96, National Archives, Suitland, Md.
23. Hellman, "The Farmers Try Group Medicine."
24. Omar Mills to John Garst, February 3, 1938, file 243, Records of the Farm Security Administration, Bancroft Library, University of California, Berkeley.
25. A. M. Simons, "Medical Service Plans of the Farm Security Administration. A Report to the Annual Conference of Secretaries and Editors of Constituent State Medical Associations," November 21, 1942. Cited in *Journal of the American Medical Association,* organizational section, 120 (1942):1315–1324. See also R. G. Leland, "Medical Care for Migratory Workers," *Journal of the American Medical Association* 114 (1940): 45–55.
26. Michael R. Grey, "Syphilis and AIDS in Belle Glade, Florida, 1942 and 1992," *Annals of Internal Medicine* 114 (1992): 329–334.
27. Baldwin, *Poverty and Politics,* 284.
28. U.S. Dept. of Agriculture, Farm Security Administration, *Annual Report of the Chief Medical Officer* (Washington, D.C.: U.S. Government Printing Office, 1939), 27. File 31-14, Mott papers.
29. U.S. Dept. of Agriculture, Farm Security Administration, *Annual Report of the Office of Medical Director, Resettlement Administration, for Fiscal Year Ending June 30, 1937,* 9–10. File 31-19, Mott papers.
30. Ibid., 5–13.
31. U.S. Congress, Senate Committee on Education and Labor. *The Experimental Health Program of the United States Department of Agriculture: A Study Made for the Subcommittee on Wartime Health and Education of the Committee on Education and Labor, United States Senate,* January 1946. This comprehensive Senate monograph was prepared by Bureau of Agricultural Economics (BAE) staff members and written by T. Wilson Longmore of the BAE. This reference is hereafter cited as BAE report.
32. Jesse B. Yaukey, *Activities of an Experimental Rural Health Program during its First Fiscal Year, 1942–43* (U.S. Dept. of Agriculture, Farm Security Administration, March 1943). File 31-2, Mott papers. See also BAE report, 37–38.
33. BAE report, 23.
34. T. Wilson Longmore and T. L. Vaughan, *Taos County Cooperative Health Association, 1942–43* (Little Rock, Ark.: U.S. Dept. of Agriculture, November 1944), 51. File 32-2, Mott papers.
35. BAE report, 47.
36. U.S. Dept. of Agriculture, Interbureau Committee to Coordinate Post-War Planning. *Experimental Rural Health Programs, Report of the Interbureau Committee for Post-War Planning* (Washington, D.C.: U.S. Government Printing Office, March 1942), 1–12. File 31-1, Mott papers.
37. BAE report, 37–38.
38. Mott and Roemer, *Rural Health and Medical Care,* 415.
39. BAE report, 75.
40. Walter Dickie, "Migrant Labor in California," *Western Journal of Medicine* 47(1938): 107.
41. *Journal of the American Medical Association,* organizational section, 108 (1936): 107b–109b.
42. Mott and Roemer, *Rural Health and Medical Care,* 406.
43. Ibid., 407.
44. Grey, "Poverty, Politics, and Health." See also Grey, "Dustbowls, Disease, and the New Deal."
45. See Starr, *The Social Transformation,* 275–277.
46. Frederick D. Mott to medical care staff, December 12, 1942. File 31-9, Mott papers.
47. BAE report, iii.
48. Ibid., xix.
49. Frederick D. Mott to health services staff, October 11, 1945. File 31-9, Mott papers.
50. Jesse B. Yaukey to Frederick D. Mott, December 22, 1946; see also U.S. Department of Agriculture, Farm Security Administration, *Annual Report of the Chief Medical Officer, Fiscal Year Ending June 30, 1945.* File 46-19, Mott papers.
51. Mott and Roemer took the position in their monograph that the loss of physician manpower and the loss of farmers to the armed services and war industries were the primary causes of the program's decline. See Mott and Roemer, *Rural Health and Medicine,* 406–407. I have argued that these forces accelerated that program's demise but that the loss of physicians' financial incentive to participate was of equal importance. See Grey, "Poverty, Politics, and Health."
52. *Journal of the American Medical Association,* organizational section, 129 (1945): 1185.
53. Frederick D. Mott to health services staff, October 11, 1945. File 31-9, Mott papers.
54. Frederick D. Mott to health services staff, November 25, 1943. File 31-9, Mott papers.
55. This issue is covered well in Starr, *The Social Transformation,* 310–334.
56. As part of my research on the FSA, I interviewed 12 individuals who participated in the FSA health programs. Many of these individuals went on to prominent positions in the public health community and participated in the social reforms that occurred in the 1960s and 1970s. Readers interested in the role that Loren Kerr played in the United Mine Workers Health and Welfare Fund are referred to Allen Derickson, "The United Mine Workers of America and the Recognition of Occupational Respiratory Diseases, 1902–1968," *American Journal of Public Health* 81 (1991):782–790. The link between FSA programs and the Office of Economic Opportunity's programs is explored in Richard Couto, *Ain't Gonna Let Nobody Push Me Around: The Pursuit of Racial Justice in the Rural South* (Philadelphia, Pa.: Temple University Press, 1991).
57. Frederick D. Mott to Jesse B. Yaukey, March 27, 1947. File 46-19, Mott papers.
58. Milton Roemer, interview with author, Seattle, Washington, October 1987.